AL/HR-TP-1995-0016

AN EVALUATIVE COMPARISON OF TECHNIQUES FOR MEASURING STUDENT SYSTEM KNOWLEDGE OF AVIONICS TROUBLESHOOTING

Nancy J. Cooke
New Mexico State University
Deparment of Psychology
Las Cruces, NM 88003

Anna L. Rowe
Rice University
Department of Psychology
Houston, TX 77251

Armstrong Laboratory
Human Resources Directorate
Brooks AFB, TX 78235-5352

Ellen P. Hall
Armstrong Laboratory
Human Resources Directorate
Brooks AFB, Texas 78235-5352

Tracy L. Halgren
New Mexico State University
Deparment of Psychology
Las Cruces, NM 88003

Ronald S. Kane
Armstrong Laboratory
Human Resources Directorate
Brooks AFB, Texas 78235-5352

Chris Kruse
33 FW, 60FS
Eglin AFB FL

Jeff A. Bauhs
New Mexico State University
Deparment of Psychology
Las Cruces, NM 88003

August 1995

Interim Technical Paper for Period December 1992-December 1993

ARMSTRONG LABORATORY

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICES

Publication of this paper does not constitute approval or disapproval of the ideas or findings. It is published in the interest of scientific and technical information (STINFO) exchange.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or an obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

Ellen Hall

ELLEN P. HALL, Ph.D.
Project Scientist
Cognitive Technologies Branch

PATRICK C. KYLLONEN, Ph.D.
Technical Director
Manpower & Personnel Research Division

GARY D. ZANK, Col, USAF
Chief
Manpower and Personnel Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1995	Final - December 1992 - December 1993

4. TITLE AND SUBTITLE

Evaluative Comparison of Techniques for Measuring Student System Knowledge of Avionics Troubleshooting

5. FUNDING NUMBERS

AFOSR Grant #F49620-90-C-09076
PE - 62205F
PR - 7719
TA - 22
WU - 03

6. AUTHOR(S)

Nancy J. Cooke, Tracy L. Halgren, Chris Kruse
Anna L. Rowe, Ronald S. Kane, Jeff A. Bauhs
Ellen P. Hall

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

New Mexico State University
Department of Psychology
Las Cruces, NM 88003

Rice University
Department of Psychology
Houston, TX 7725

33 FW, 60FS
Eglin AFB FL

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Manpower & Personnel Research Division
7909 Lindbergh Drive
Brooks AFB, Texas 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/HR-TP-1995-0016

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Ellen P. Hall, (210) 536-3570

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Intelligent tutors have the potential to enhance training in avionics troubleshooting by giving students more experience with specific problems. Part of the success of these tutors will depend on their ability to assess and diagnose the student's knowledge in order to direct pedagogical interventions. The broad goal of the research program described here is to develop a methodology for assessment and diagnosis of student knowledge in the domain of avionics troubleshooting. One step towards this involves the identification and evaluation of techniques for measuring students' system knowledge in the domain of avionics troubleshooting. This step is the focus of the research described in this report. Thus far, the results of this research program have (1) developed a means of identifying meaningful action patterns in on-line student-tutor interactions, (2) identified techniques for measuring students' system knowledge which predict troubleshooting performance, and (3) described how these techniques contribute to on-line student-tutor assessment.

14. SUBJECT TERMS

Intelligent tutoring
Knowledge Assessment
System Knowledge
Student Modelling

15. NUMBER OF PAGES

77

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

This research was supported by the Air Force Office of Scientific Research (AFOSR) Research Initiation Program - Grant No. F49620-90-C-09076, and the Air Force Armstrong Laboratory, Human Resources Directorate. Portions of this research were conducted at Rice University and New Mexico State University. The study was conducted as part of the second author's doctoral dissertation. All of the opinions expressed in this paper are those of the authors.

We would like to thank the aircraft maintenance technicians assigned to Air Combat Command who contributed to the study. This work could not have been accomplished without their support.

AN EVALUATIVE COMPARISON OF TECHNIQUES FOR MEASURING STUDENT SYSTEM KNOWLEDGE OF AVIONICS TROUBLESHOOTING

INTRODUCTION

As tasks become more cognitively complex and demand more specialized skill, training issues are increasingly critical. The domain of avionics troubleshooting is a good example of such a task. Recent research (e.g., Nichols, Pokorny, Jones, Gott, & Alley, 1989) suggests that computerized intelligent tutoring systems may successfully supplement traditional training in complex tasks such as avionics troubleshooting where problems are ill-defined. Intelligent tutoring systems enable individuals to spend time learning a skill in a one-on-one environment in which a computer takes on the role of a human tutor. One goal of intelligent tutoring systems is to incorporate more individualized instruction based on a detailed assessment of student knowledge and diagnosis of cognitive strengths and weaknesses. Instructional intervention can then be directed at these strengths and weaknesses. The purpose of the work described here is to develop a methodology for the assessment and diagnosis of student knowledge in the domain of avionics troubleshooting.

The problem of assessment and diagnosis for intelligent tutoring systems has been approached in a number of ways. Many approaches focus on student actions to assess and diagnose student knowedge. For example, student actions may be examined for the identification of incorrect actions (e.g., Burton, 1982; Gitomer, 1992; Pokorny & Gott, 1994; Stevens, Collins, & Goldin, 1979), or student actions may be compared to a logically created ideal action set (e.g., Anderson, Boyle, & Reiser, 1985). One problem with these approaches is that the action data are impoverished relative to the much richer data obtained from cognitive methods such as verbal reports and structural analyses. These richer methods go beyond the student's actions to reveal the cognitive processes and knowledge underlying those actions. Such information is particularly critical for the identification of the knowledge base in complex, ill-specified domains like avionics troubleshooting. Furthermore, such information is important in the assessment and diagnosis of students' system knowledge. These issues are discussed in detail below.

Unlike the typical intelligent tutoring domains of algebra, geometry, and computer programming, avionics troubleshooting is ill-specified and is not associated with a well-documented body of knowledge. In these types of domains, before ideal models or buggy rules can be specified, it is first necessary to identify the specific knowledge necessary to perform the task. One cognitive task analysis (Hall, Gott, & Pokorny, 1994) of the avionics troubleshooting domain has indicated that three types of knowledge are relevant for successful troubleshooting performance: (1) system (or how it works) knowledge, (2) strategic (or how-to-decide-what-to-

do-and-when) knowledge, and (3) procedural (or how-to-do-it) knowledge (Gott, 1989). Evidence exists that suggests that system knowledge or mental model knowledge may be the most critical of these three (Gitomer, 1984; Glaser et al., 1985; Hall, Gott, & Pokorny, 1994; Rasmussen & Jensen, 1974), although this point is not without controversy (Kieras, 1988; Rouse & Morris, 1986). Thus, the methodology developed in this research program focuses on the assessment and diagnosis of system knowledge.

Observations of student actions can reveal information about procedural and strategic knowledge; however, these observations are less likely to disclose system knowledge. Instead, relatively rich cognitive techniques have been used in the past to elicit this type of mental model knowledge. These techniques can be classified into four categories: (1) accuracy and time measures, (2) interviews, (3) process tracing/protocol analysis, and (4) structural analysis (Cooke & Rowe, 1993). Measurement methods drawn from each of these categories have advantages and disadvantages (Cooke, in press), and no one method of measuring mental models has received universal acceptance. The different measurement approaches may each provide different sorts of information and have seldom been evaluated in terms of their respective reliabilities and validities. In short, the selection of a single optimal method for on-line student assessment of system knowledge is an uncertain enterprise at best. In this paper a pragmatic approach is taken in which optimal methods are minimally assumed to elicit system knowledge that is relevant to task performance.

Although rich cognitive methods seem better suited for measuring system knowledge than the action-oriented assessment approaches described earlier, they are not well-suited for on-line measurement. Instead, these methods typically involve the collection of "extra" data (e.g., verbal reports, similarity ratings) not typically collected during interactions with the tutor. Consequently, the use of these methods would entail interruption of the tutoring process to collect data in a task that would likely seem artificial to the student. What is needed is not only a sound method for measuring system knowledge, but one that can provide rich representations of this knowledge from student actions derived on-line. Such a methodology is the overall goal of our research program. The goal is to be able to map student actions (both errorful and correct) collected on-line onto a rich representation of student system knowledge. This representation can then be used to assess and diagnose student system knowledge and identify targets for intervention.

The Broad Plan: Mapping Student Actions onto System Knowledge

The overall goal of our research program involves making detailed inferences about a student's system knowledge from that student's actions. Four steps make such an inference possible (Cooke & Rowe, 1993). The first step involves working backwards from the goal state--system knowledge, to the initial state--student actions. Interviews, process tracing, and structural

analytic methods offer rich representations of system knowledge. However, it is necessary to know which of these methods provides the best measure of system knowledge in the domain of avionics troubleshooting (see Figure 1, Step 1). Therefore, the first step in reaching the overall goal involves identifying a valid method for eliciting and representing system knowledge required for avionics troubleshooting. Assuming that system knowledge is critical for performance, then a valid method of measuring this knowledge should reveal differences among subjects that correspond to performance differences.

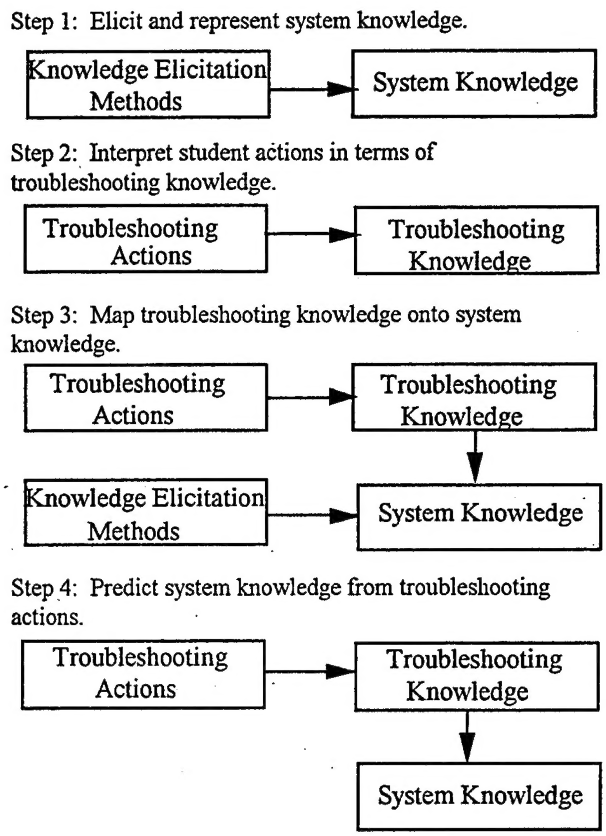

Figure 1. Steps involved in mapping student actions onto system knowledge.

As mentioned above these techniques require data collected off-line. Therefore, the next step involves determining how to derive this type of data from on-line interactions with the tutor. What is needed is a method designed to make use of action data collected on-line to derive representations of system knowledge. In other words a method is needed for identifying general relationships between student action patterns and patterns of system knowledge derived off-line, so that later predictions can be made about system knowledge based on student actions. Thus, the identification of action patterns and the evaluation of the meaningfulness of these patterns is the second step (see Figure 1, Step 2). Along this line of reasoning, the third step entails mapping

these action patterns onto patterns of system knowledge (see Figure 1, Step 3). Here, the goal is the identification of patterns of actions that correspond to distinct representations of system knowledge. Of course this step requires the elicitation of both actions and system knowledge from the same subjects. Assuming that Step 1 results in meaningful representations of system knowledge and assuming that system knowledge underlies actions (at least partially), then some correspondence should emerge. Finally, if this correspondence does emerge, then it would be possible to make predictions about system knowledge from troubleshooting actions collected on-line, thereby eliminating the extra data collection step (see Figure 1, Step 4). The four steps represented in Figure 1 comprise the long-term plan associated with the development of a new approach for assessing and diagnosing student system knowledge.

The contributions of this broad research plan are as follows:

1. A procedure for on-line assessment and diagnosis of student's system knowledge which involves mapping action patterns onto deficits or proficiencies in system knowledge.
2. A procedure which summarizes actions (errorful and correct) in terms of a rich representation of student knowledge that lends itself to qualitative analysis useful for diagnosis and intervention.
3. An assessment and diagnosis procedure that targets the complex domain of avionics troubleshooting.
4. A methodology that can be applied to the problem of eliciting knowledge from subject matter experts for tutor development.
5. A general test of the assumption that system knowledge underlies troubleshooting actions.

Previous Work

Research pertinent to the second step (identifying meaningful action patterns) was completed by Cooke and Rowe (1993). This particular step was carried out first because the data required were already available. Cooke and Rowe examined verbal troubleshooting action data collected by Nichols et al. (1989). The Pathfinder network scaling procedure (Schvaneveldt, 1990) was used to summarize the actions executed by subjects on a set of verbal troubleshooting tests. The results are promising in that they indicate that meaningful action patterns can be identified using this procedure. Specifically, the network patterns were predictive of troubleshooting performance (r (22)= .57), where performance was defined by a previous scoring method (Pokorny & Gott, 1994). Furthermore, using the network patterns, Cooke and Rowe were able to differentiate tutor vs. no tutor groups in the Nichols et al. (1989) study. Subjects who were trained on the tutor had networks that were more similar to the ideal network than did subjects in the control condition. This finding indicates that subjects who were trained on the tutor learned more expert-like action sequences than did subjects who were not. Finally, the

networks revealed qualitative differences in the action sequences of high and low performers which were suggestive of potential intervention points. Thus, Cooke and Rowe identified a technique that captures meaningful action patterns during troubleshooting.

Current Work

The purpose of the research described in this report is the evaluation of different measures of system knowledge, primarily in terms of their ability to elicit system knowledge or mental model knowledge that is predictive of troubleshooting performance (i.e., the first step). The current research differs from other studies which have used multiple methods for measuring mental models (e.g., Gray, 1990; Gitomer, 1984; McCloskey, 1983) in that the primary interest here is a comparison of the measurement techniques themselves. The primary interest in previous research endeavors employing multiple measurement techniques has been the assessment of underlying cognitive characteristics (including, but not limited to, mental models), using converging operations. This research also addressed several issues which have been raised in the conduct of previous work in the avionics troubleshooting domain, including the identification of a performance measure. This issue and others were addressed by (1) utilizing measurement techniques within a specific troubleshooting context, (2) employing cognitively rich mental model measurement techniques that have shown promising results in other domains, including pairwise comparisons, Pathfinder, and think aloud reports, (3) using a performance-based criterion (i.e., verbal troubleshooting score), and (4) measuring performance on a continuous scale rather than a dichotomous scale in attempt to increase measurement sensitivity.

The research consists of two general phases: a problem selection phase and a system knowledge measurement technique comparison phase. During the problem selection phase, a moderately difficult troubleshooting problem was selected. Problem selection was vital because all experimental materials revolved around one problem. During the measurement technique comparison phase, subjects completed each of four system knowledge measures. These measures were drawn from each of the major categories outlined above, excepting the time and accuracy measures category (i.e., interviews, process tracing/protocol analysis, and structural analysis). Specifically, subjects' mental models of an avionics system were measured using: a laddering structured interview, concept relatedness ratings, a diagramming structured interview, and think aloud while troubleshooting. All mental model measures took place within the context of a specific troubleshooting problem. In addition to completing the mental model measures, each subject worked to verbally troubleshoot the problem. Relating performance on each of the mental model measures to troubleshooting performance should offer insight into the strengths and weaknesses of each of the measures for accessing knowledge pertinent for performance.

METHOD

Problem Selection

All experimental materials were developed in the context of a particular troubleshooting problem in the F-15 flightline avionics communications, navigation, and electronic warfare systems (or C Shop) career field of the U.S. Air Force. The procedure used to select this problem was designed for the selection of a moderately difficult troubleshooting problem, a problem presumably requiring the invocation of a mental model for successfull troubleshooting. Such a problem should distinguish expert from novice technicians. Much of the data used in the selection of this problem were gathered during Stages I-VIII of a PARI cognitive task analysis (Hall, Gott, & Pokorny, 1994) conducted by an Air Force research team in the C Shop career field (Hall, Pokorny, & Kane, 1994) at Eglin Air Force Base.

PARI (Precursor, Action, Result, Interpretation) is a cognitive task analysis methodology used by the Air Force as an integrated skill analysis/instructional development tool. The PARI data collection procedure consists of nine stages. In general, the first four stages serve to identify a sample of subject matter experts. These experts then assist the research team in identifying the general problem solving tasks encountered in the career field and the cognitive skills associated with successfully solving these tasks. The final five stages of PARI involve the development of problem-solving scenarios and the collection of problem-solving interview data from experts and novices as well as a set of follow-up reviews of the data. Hall, Pokorny, and Kane's (1994) PARI data were used in the selection of the moderately difficult troubleshooting problem. Data from Stage IX of this PARI analysis were not pertinent in the selection of a moderately difficult troubleshooting problem and thus were not utilized. The following paragraphs offer only a brief description of the PARI methodology; a more complete account of PARI can be found in Hall, Gott, and Pokorny (1994).

PARI-Stage I. The first stage of PARI is designed to identify subject matter experts who then go on to participate in the remaining PARI stages. In order to identify C Shop experts, the Air Force research team conducted individual discussion sessions with technicians who had been identified as the most highly skilled by shop supervisors and were available for participation in the discussions. During the discussion session, the technician was asked to iteratively break down F-15 avionics equipment systems in terms of their component parts. First, the technician identified the subsystems of a particular system (e.g., the Radar Warning Receiver or RWR system is part of the Tactical Electronic Warfare System or TEWS). The technician then broke the subsystems down to the component level, identifying the function of each component as it was named. This break-down continued until the technician believed that the components at the lowest level could not be further subdivided. In addition to iteratively breaking down the equipment systems, the technician addressed job training problems associated with the particular system.

Based on these discussions, the researchers, as a group, determined which of the sampled technicians qualified as experts. This determination was based on three aspects of each technician's discussion: (1) the quality of the verbalized equipment representations, (2) the identification of specific equipment component relations, and (3) the level of clarity in the technician's equipment descriptions. Selection was also based on availability to participate in the PARI sessions. Following the application of this process, the research team designated two of the sampled technicians as subject matter experts: an Air Force Technical Sergeant and an Air Force civil servant. These experts assisted the research team through the remaining stages of the cognitive task analysis.

PARI-Stage II. The second stage of PARI is designed to establish the training foci associated with the job in question. (PARI was developed to assist in the development of training that targets complex problems.) To form the C Shop training foci, the two experts worked to list and discuss the maintenance tasks (i.e., troubleshooting problems) they felt were difficult. These discussions were facilitated by an exhaustive listing of maintenance tasks for that career field provided by the Air Force Specialty Training Standards. The Air Force research team used two related criteria to classify tasks as cognitively complex: the degree of decision-making required in performing the task and the stability of the task (or system) environment in which problem solving occurs. Tasks were considered cognitively complex if they required decision making (i.e., a procedure specifying step-by-step actions for solving the problem is not available) and if they occurred in an unstable environment (i.e., many factors must be considered in making decisions). Tasks meeting these two criteria were subsequently used by the research team to facilitate discussion with the two experts. The experts were asked to identify maintenance tasks associated with their jobs which are cognitively demanding and to discuss their reasoning for this assertion. The team and the experts then decided together whether or not to categorize the task as cognitively complex. Thus, the research team and the experts worked together to identify the cognitively complex tasks associated with the C Shop.

PARI-Stage III. The purpose of the third stage of PARI is the generation and consolidation of the problem types encountered by technicians working on the job. Using the cognitively complex tasks identified in Stage II, the experts worked to generate an exhaustive list of the equipment malfunctions (and their causes) that could initiate troubleshooting for these tasks. The experts independently specified fault instances in cause and effect language (e.g., "bad stimulus routing caused by a stuck relay") for each of the defined tasks. The experts then worked together to consolidate the identified system causes and effects into meaningful categories (e.g., wiring faults). Fault instances were grouped together if they demanded similar knowledge and skills for solution. This grouping resulted in the following typology of problems that could initiate

troubleshooting in the C Shop: set up procedure faults, switchology faults, cable faults, wiring faults, and electrical/component faults.

PARI-Stage IV. The fourth stage of PARI involves the development of representative troubleshooting problems for each of the problem categories specified in the problem typology. The typology serves to ensure that representative examples of all problem types are generated. The experts individually developed a representative troubleshooting problem for each of the five problem categories. For each of these problems the experts: (1) developed an overview or problem description that listed the fault location and the symptoms associated with the fault, (2) generated a problem statement that listed the system conditions and symptoms for presentation to other individuals for troubleshooting, and (3) listed the supporting technical documentation (e.g., test procedures, schematics) that would be required by others troubleshooting the problem. At the conclusion of PARI-Stage IV, each expert had designed one troubleshooting problem for each problem category.

PARI-Stage V. The purpose of the fifth stage of PARI is the anticipation of the supporting information (e.g., Technical Orders or T.O.s) technicians would require to solve the developed problems. In addition, this stage also provides an opportunity for the experts to specify exactly how the various pieces of equipment would function under the faulty conditions. To obtain these sorts of information, the experts worked individually to generate their own solutions to the problems they had developed. The PARI problem-solving structure was used to guide solution generation: Actions, Precursors to actions, Results, and Interpretations of results were elicited. Following the recording of the initial solution, five "rehashes" were conducted in which the expert worked to (1) verify the initial solution, (2) generate alternative results and result interpretations for each step in the solution, (3) identify and evaluate alternative actions for each step in the solution, (4) name alternative appropriate equipment targets (precursors), given the previously executed steps, and (5) group the actions that seem to go together and to explain the basis for these groupings. At the conclusion of PARI-Stage V, the experts had each generated one solution for each of their respective problems.

PARI-Stage VI. During the sixth stage of PARI, experts naive to the problems are asked to generate solutions to those problems. Thus, two additional technicians were asked to produce solutions to the generated problems. These technicians had been identified by the two experts as skilled technicians, and both were at the 7-Skill Level. The Air Force uses a four level classification system to designate skill level: all technicians start at the 3-level and move from the 5- to the 7-level after passing certain training criteria (e.g., demonstrating proficiency running an operational checkout of a particular equipment system). Technicians at the 7-level have reached the highest level of technical proficiency. They then move on to the 9-level which designates

them as qualified on management-related tasks. The 9-level designation is attainable only by Senior and Chief Master Sergeants, who no longer have hands-on maintenance responsibilities.

The two technicians worked individually with each of the experts to solve the generated troubleshooting problems. The experts presented only the problems they had developed. The expert began by presenting the problem statement to the technician (e.g., "During debrief, the crew chief reports an ASP 44"). The technician then worked to isolate the fault and repair the equipment through a series of iterative action-result steps. In each step the technician specified an action and the reason for taking that particular action. The expert responded by informing the technician of the action's effect on the equipment, and requested the technician's inference concerning equipment operation based on that result. If the technician strayed from the active path during troubleshooting and did not return in a timely manner, the expert presenter provided coaching. The action-result cycle continued until the problem was solved. This procedure was followed for each problem in the problem set. See Table 1 for a sample of three steps drawn from a solution given for the ASP 44 problem.

Table 1
Three steps drawn from a solution given for the ASP44 problem. (Problem statement: In debrief, the crew chief reports an ASP44.) Note: P = Precursor, A = Action, R = Result, and I = Interpretation.

PARI	Technician Response
Step 1	
P:	Sometimes ASP44 will come down in flight. Also, there was no BIT light and the pilot didn't report any discrepancies.
A:	Run up aircraft, and see if ASP44 clears.
R:	ASP44 does not clear.
I:	Most likely the LRU6 is bad.
Step 2	
P:	LRU6 is probably bad.
A:	Remove and replace the LRU6.
R:	ASP 44 latches. No BIT light.
I:	Something is causing the ASP44 to latch.
Step 3	
P:	Fault indications go through the LRU3 before being sent to the ASP.
A:	Remove and replace the LRU3.
R:	ASP44 does not clear.
I:	Must have a bad wire somewhere.

PARI-Stage VII. During the seventh stage of PARI, the expert problem developers review the problem set to determine the adequacy with which the developed problems assess the cognitive skills and knowledge required for skilled performance. The two expert developers reviewed the problems and ascertained that they were representative of problems encountered in the actual job environment. In addition, they determined that solutions to the problems required the types of cognitive skills and knowledge associated with skilled performance. In addition to making these judgments, one of the experts rank-ordered the troubleshooting problems in terms of their difficulty (see Table 2). His rank-ordering was based on his observations of the technicians troubleshooting the problems, in addition to his 12 years of experience in the C Shop. The two additional expert technicians who had worked to solve the developed problems were unable to solve the following three problems without coaching from the expert presenter: ICS Frequency Holes, ASP44, and CMD Safety Switch. Thus, these three problems were considered more than moderately difficult and were dropped from consideration, leaving the RWR-RF Loss problem as the most likely candidate for problem selection.

Table 2
Rank ordering of troubleshooting problems from most (1) to least (10) difficult offered by one of the expert problem developers.

Troubleshooting Problem Ranking
1. ICS Frequency Holes
2. RWR--RF Loss
3. ASP 44
4. CMD Safety Switch to Scope
5. RWR Wiring--Broken Wire
6. AAI--Cable at Radar
7. IFF Seat Switch
8. KY-58 Relay Panel
9. 51 & 52 Switches--AAI
10. Have Quick Word of Day

PARI-Stage VIII. During the eighth stage of PARI, less-skilled technicians are observed as they work to generate solutions for the developed problems. Thus, intermediate and novice technicians were asked to troubleshoot the remaining problems, using the PARI problem-solving structure described above. One of the expert problem developers presented the problems he had developed to the technicians. The remaining problems were presented by an expert from the Air Force research team (the expert problem developer was unavailable). Observations of the

troubleshooting behavior of these technicians revealed that novice technicians could quickly isolate the fault associated with RWR--RF Loss problem. In addition, they could easily solve the problem if they had previous experience with a specific cable measurement tool, although this tool was only just being introduced to the C Shop and was not in common use. Because the less-skilled technicians could quickly and easily isolate the fault, the RWR--RF Loss problem was deemed inappropriate for selection as the moderately difficult troubleshooting problem. On the other hand, observations of technicians solving the RWR Wiring--Broken Wire problem indicated that this problem revealed performance differences across technicians of different skill levels. Thus, this problem was selected for use as the troubleshooting problem in this study.

The RWR Wiring--Broken Wire Problem

The system important in troubleshooting the RWR Wiring--Broken Wire problem is the Radar Warning Receiver (RWR) system. The RWR is part of the F-15 Tactical Electronic Warfare System (TEWS) and is designed to detect, analyze, and identify threat radar signals. The RWR also controls countermeasure responses to those threats (e.g., the release of chaff). The RWR system consists of eleven components: six line replaceable units (LRUs) and five antennas. The fault in this problem is a shorted video cable between two of the LRUs. Details regarding the RWR system and the selected troubleshooting problem can be found in Appendix A.

Subjects

Subjects were nineteen technicians in the F-15 flightline avionics (C Shop) career field of the Air Force. Each subject had been through a technical training school designed to prepare them for their specialty and had received subsequent on-the-job training. Technicians working in the C Shop are responsible for the identification, isolation, and repair of airborne avionics equipment systems, including the RWR and Identification--Friend or Foe (IFF) subsystems. The technicians participating in this study all worked directly with the equipment and were selected to achieve a range of proficiency. Thus, six 3-levels, eight 5-levels, and five 7-levels participated. All participating technicians but one were male.

Materials and Procedure

Technicians' mental models of an avionics system (i.e., the RWR) were measured using four techniques: a laddering structured interview, concept relatedness ratings, a diagramming structured interview, and a think-aloud troubleshooting protocol. Technicians' knowledge was also measured using a retrospective interview, but these data were collected for other purposes and will not be discussed here. All mental model measures took place within the context of the RWR Wiring--Broken Wire troubleshooting problem. In addition to completing all of the mental model measures, each technician worked to verbally troubleshoot this problem. The procedural steps are outlined below.

Laddering Structured Interview. Upon arrival to the testing session, the technician completed the laddering structured interview. This interview consisted of four steps. The troubleshooting problem statement was presented to the technician: "In debrief, the pilot reports that the RWR is inoperative, the BIT (built-in-test) light is on, and the TEWS display is blank." The first step of the interview consisted of asking the technician to identify the major system important in troubleshooting this problem. In Step 2 the technician was asked to name the major components of the identified system, in the context of the troubleshooting problem. In the third step of the interview the technician was asked to name all of the major components of the identified system, regardless of problem context. In the fourth and final step the technician was asked to name the major systems with which the identified system interfaced, if any. Throughout these steps an index card with the identified component or system name written on it was prepared and placed before the technician, according to the arrangement specified by the technician. The index cards only served as memory aids for the technician and were discarded after the identified systems and components had been recorded.

Relatedness Ratings. Following the laddering structured interview, the technician completed two sets of ratings on the eleven RWR system components. Both sets of ratings were completed on a Macintosh computer, using a Hypercard program to collect the data. The technician first completed familiarity ratings by using a mouse to point and click on one of six sections on a bar, with the endpoints labeled *unfamiliar* and *highly familiar*. The components were presented in a random order.

The technician then completed relatedness ratings on all pairs of the same eleven components. The technician was told to make all relatedness ratings within the context of the RWR Wiring--Broken Wire troubleshooting problem. The technician was told to rate the items in terms of their functional relations. The technician was told that although two items can be functionally related in a number of different ways (e.g., by information flow through the system, by performing similar function), ratings should be made based on the first general impression of functional relatedness of the concepts, within the context of the troubleshooting problem.

The technician rated the relatedness of the component pairs by using a mouse to point and click on one of five sections on a bar, with the endpoints labeled *slightly related* and *highly related*. If the technician wished to rate the component pairs as unrelated, a button labeled *unrelated* was available. The troubleshooting problem statement was available for review at the top of the computer screen throughout the relatedness ratings task. Presentation of pairs was randomized across subjects, and order of items within pairs was counterbalanced.

Diagramming. After completing the ratings, the technician completed a diagramming task, using the component set just rated. Randomly ordered index cards with the name of an RWR component printed on each were given to the technician, as was a set of directional and bi-

directional white arrows. The technician was instructed to arrange and connect the components in a manner representing the actual function of the RWR system, in general. The technician specified directionality of relations with the unidirectional and bi-directional arrows. To illustrate the use of the arrows for representing functional relations, the technician was given an example diagram with an accompanying explanation from the domain of automobile engines (see Figure 2).

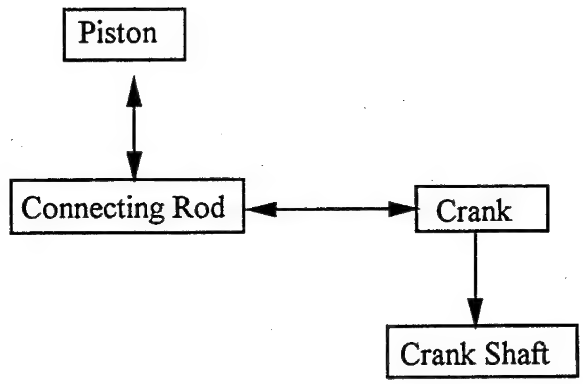

Figure 2. Example functional diagram from the domain of automobile engines.

After the technician completed the functional diagram representing the RWR system at a general level, the troubleshooting problem statement was re-presented. The technician was then given a set of directional and bi-directional yellow arrows and was asked to use these yellow arrows to designate those components and/or connections most important in troubleshooting the problem. Finally, the technician was asked to explain, in his/her own words, both diagrams. These explanation data were collected to aid the examiner's understanding of the generated diagrams, and no further analyses were conducted on these data.

Think Aloud. The technician then proceeded to the think aloud while troubleshooting portion of the experiment. The technician was told that s/he would be verbally troubleshooting the problem used in each of the previous tasks, the goal being to isolate the fault and repair the equipment. The technician was instructed to think aloud continuously while working to solve the problem, verbally expressing all thoughts. Two practice think-aloud problems were then reviewed with the technician to ensure that the technician understood what was meant by thinking aloud (i.e., what is the result of multiplying 24 by 6, and how many windows are there in your house?). If the technician had difficulties during these practice problems (e.g., did not speak), s/he was guided to think aloud. After successfully completing these practice think-aloud problems, the verbal troubleshooting session began.

The examiner re-presented the troubleshooting problem statement. All technical materials necessary for troubleshooting the problem were available. These materials included the C Shop Job Guide (J.G.) and the T.O. which contains fault isolation trees and schematic diagrams of the

RWR system. The technician was instructed that the goal was to isolate the fault and repair the equipment through a series of iterative action-result steps. In addition, the technician was reminded to verbally express all thoughts while working to solve the problem. An expert assisted with this task by "simulating" the equipment for the technician. Specifically, the expert provided the technician with results for all specified actions. The technician began by specifying the first action s/he would take in troubleshooting the problem (e.g., check the Avionics Status Panel or ASP). The expert responded by informing the technician of the action's result (e.g., ASPs 5 and 49 latched). The technician then specified the next action, the expert gave the corresponding result, and so on. This action-result cycle continued until the problem was solved, the 45-minute time limit expired, or the technician gave up. All of the technician's, the expert's, and the examiner's responses were recorded with an audio tape recorder.

After completing this task, the examiner and the technician reviewed the protocol action by action. The technician was asked to provide a retrospective report of: (1) why each action was taken, and (2) the information provided by the corresponding result of each action, in terms of the equipment in question. This retrospective interview was conducted for other purposes and will not be further discussed here.

Questionnaires. Following the troubleshooting action review, the technician completed two questionnaires. First, the technician completed a Likert-style questionnaire designed to allow an evaluation of the mental model measurement techniques (see Appendix B). Specifically, the technician used a 6-point scale to rate each of the techniques on the following dimensions: difficulty (*difficult--easy*), similarity to actual troubleshooting in the shop (*different--similar*), range of responses available (*restricted--broad*), realism relative to actual troubleshooting in the shop (*artificial--realistic*), and usefulness for measuring system knowledge (*useless--useful*). Space was available for additional written comments.

The technician was then given a questionnaire designed to allow a comparison of the measurement techniques (see Appendix C). All pairs of tasks (measurement techniques and troubleshooting task) were presented on this questionnaire. The technician was asked to make pairwise judgments of the tasks by circling the task in the presented pair which best measured knowledge needed for actual troubleshooting of the RWR Wiring--Broken Wiring problem in the shop. Immediately after circling one of the tasks in the presented pair, the technician used an 8-point scale to rate the similarity of the knowledge measured by the circled task to knowledge needed for the actual troubleshooting of a problem in the shop. The endpoints of the scale were labeled *Not at all similar* and *Extremely similar*. Upon completion of this questionnaire, the technician was debriefed and excused. Separate from the testing session, a supervisor rating questionnaire was completed for each technician by his/her respective supervisor (see Appendix D).

DATA ANALYSIS

Overview of Analyses

The goal of this study was to evaluate the various mental model measures in terms of their abilities to predict troubleshooting performance. The "bottom line" in real-world troubleshooting situations is performance. Thus, comparisons of the results obtained from the measurement techniques with the results from the troubleshooting task should provide a pragmatic means of assessing the validity of the techniques. This approach assumes that a high-quality mental model should be associated with high-quality troubleshooting performance, and two assessments are required: an evaluation of troubleshooting performance and an assessment of the technicians' mental model knowledge.

Troubleshooting Performance

To obtain a troubleshooting performance measure, two subject matter experts independently scored the technicians' troubleshooting action protocols, using a modified Q sort. These experts had participated in the troubleshooting problem development stage. In addition, one of the experts (Expert B) assisted the examiner in problem presentation when the technicians worked to troubleshoot the problem. The protocols given to the experts for scoring contained the actions verbalized by individual technicians, along with the corresponding results of those actions. The experts were instructed to score the protocols based only on the listed actions and results. The experts read each of the protocols and rank ordered them according to the troubleshooting proficiency displayed. The experts then re-read the protocols and assigned a score to each protocol, using a 100-point scale where 100 indicated correct and efficient fault isolation. Figure 3 represents the troubleshooting performance scores awarded to the 19 subjects by Experts A and B.

An examination of the Figure 3 reveals that the two experts used the scale differently in scoring the troubleshooting protocols. Expert A's scores ranged from 0 to 100, whereas Expert B's scores ranged from 65 to 100. However, the scores awarded by the two experts were significantly correlated[1] $r\ (17) = .883$, $p < .0001$, as were the rank-orderings given by the experts, Spearman's $r\ (17) = .847$, $p < .0005$. Thus, a single performance score was created for each technician by averaging the troubleshooting scores given by the two experts. This score was then used as the performance score in all subsequent analyses. Four high performers emerged from this analysis who had performance scores one standard deviation greater than the mean (i.e., all scores 95 or above).

[1] All correlations reported in this paper are Pearson product moment correlations unless otherwise specified. In addition, because correlations between performance and knowledge were expected in particular directions, they were tested for significance using one-tailed probabilities.

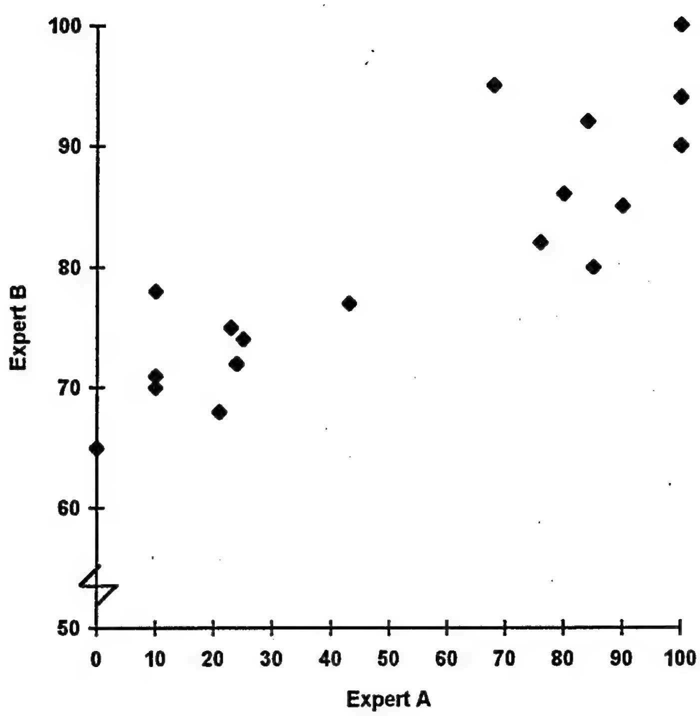

Figure 3. Scattergram of troubleshooting performance scores awarded to the nineteen subjects by Experts A and B.

Mental Model Knowledge

Whereas performance is generally indicated in terms of reference to some ideal or perfect score, the results of each of the knowledge measures offer only information regarding the content of technicians' knowledge. Therefore, what is needed is an assessment of the technicians' knowledge. Such an assessment requires an ideal or "standard" mental model. A standard was created for each mental model measure by combining the data from the four technicians who scored the highest on the performance measure, the verbal troubleshooting task. The group's combined results comprised the standard for that measure. In cases in which there was little agreement across high performers, the group was limited to those who agreed. The quality of each technicians' mental model knowledge was then assessed in terms of the overlap between this standard and the technicians' response to the measure. The resulting knowledge assessment scores were then correlated with the performance score. This comparison provides a practical means of assessing the mental model measures.

RESULTS AND DISCUSSION

Laddering Structured Interview

To determine the level of inter-subject agreement among the four high performers, the proportion of shared items across lists for each pair of high performers was calculated for each step in the interview (see Table 3). That is, the ratio of shared items to the total number of different items listed was calculated for each step for all pairs of high performers. The resulting proportions indicate that the high performers agreed on the important components or systems for each step, particularly Steps 1 and 2. All high performers listed the RWR as the system important in troubleshooting the problem (Step 1). In addition, for Step 2 each of the high performers named at least three of the same components (i.e., the LRU2, LRU3, and LRU9) as important components in troubleshooting this problem.

Table 3

The proportion of shared list items for pairs of high performers for each interview step.

		Technician Number		
	Step	6	8	14
5	Step 1	1.0	1.0	1.0
	Step 2	1.0	.75	.50
	Step 3	.67	.63	.71
	Step 4	.50	.67	1.0
6	Step 1		1.0	1.0
	Step 2		.67	.50
	Step 3		.63	.60
	Step 4		.40	.50
8	Step 1			1.0
	Step 2			.50
	Step 3			.67
	Step 4			.67

A standard component/system list was created from the lists of these four high performers for each of the four steps (see Table 4). Items named by at least one of the high performers were included in the list. Knowledge indices for the 15 remaining technicians were then calculated in terms of the proportion of items shared with the standard list associated with each step of the

interview. The resulting knowledge indices for each step were correlated with the troubleshooting performance score, excepting Step 1 (see Table 5). At Step 1, all technicians named the RWR as the system important in troubleshooting the problem, indicating that all sampled technicians correctly interpreted the presented symptoms at a coarse level. Correlations conducted on the remaining three steps indicated that naming components important for troubleshooting the problem (Step 2) was predictive of troubleshooting performance, $\underline{r}$ (13) = .542, $\underline{p} < .025$. This positive relationship indicates that good troubleshooters agreed with the high performers on the components important for troubleshooting the problem, whereas poor troubleshooters did not. The data resulting from Steps 3 and 4 (i.e., name all components regardless of context, and name interfacing systems) were not predictive of troubleshooting performance (see Table 5).

Table 4

Standard lists for each interview step. Note: The Step 3 list includes items named by individual technicians as major components of the identified system, regardless of problem context, in addition to those items named by that technician in Step 2.

Laddering Interview Steps			
Step 1	Step 2	Step 3	Step 4
RWR	LRU2	LRU6	EWWS
	LRU3	LRU10	ICMS
	LRU6	LRU11	Blanker
	LRU9	Left Wing Antenna	CC
	LRU10	Right Wing Antenna	CMD
	Aircraft Wiring	Left Fin Antenna	
		Right Fin Antenna	
		Low Band Antenna	
		Antenna Cables	
		ASP	

Note, the multistage Bonferroni procedure (Larzelere & Mulaik, 1977) was used to control for inflation of the Type I error rate within tests conducted on the laddering interview steps measure. Alpha family-wise (α_{FW}) was set at .10. Only correlations tested at the first stage of the procedure were significant, with alpha test-wise (α_{TW}) = .03. Likewise, in the analyses that follow each mental model measure was treated as an independent measure, and this same procedure for protecting against inflation of the Type I error rate was followed for each measure.

Only correlations tested at the first stage of the procedure for each measure were significant, thus only one α_{TW} value is reported.

Table 5
Correlations between troubleshooting performance and knowledge indices: (1) the proportion of shared items with the standard list and (2) errors of commission for Steps 2-4 of the laddering interview.

	Proportion of Shared Items	Errors of Commission
Step 2	.542	.505
Step 3	-.128	-.326
Step 4	-.200	-.296

In addition to examining the proportion of shared components/systems, the number of items listed by technicians, but not by the high performers, was calculated for each interview step for each technician. These errors of commission made up a second set of knowledge indices which were then correlated with troubleshooting performance. Again, data from Step 1 were not included in these correlations because all technicians named the RWR as the important system in troubleshooting the problem, and no errors of commission were made. Of the remaining three steps, only errors of commission occurring during Step 2 of the interview were predictive of troubleshooting performance, $\underline{r}$ (13) = .505, $\underline{p} < .05$. This marginally positive correlation indicates that technicians who identified extra RWR system components important in troubleshooting the problem (not included in the standard list) were better troubleshooters than were technicians who did not. The correlations resulting from the remaining steps were not significant (See Table 5, $\alpha_{TW} = .03$).

In general, the second step of the laddering structured interview is significantly related to troubleshooting performance. Those technicians who listed more components that were shared with the standard list performed the troubleshooting task better than those who listed fewer standard components. Unexpectedly, those technicians who also listed more components <u>not</u> on the standard list were better troubleshooters than those who listed fewer. This latter result, although unexpected, corroborates earlier findings in this area. Specifically, Cooke and Rowe (1993) found that as students gained troubleshooting experience, they tended to execute a greater number of actions (and even more than high performers), although they did not seem to know when the actions should be applied. In a similar way, the best troubleshooters in this study believed that many components were relevant for troubleshooting the problem, including those which are actually relevant. Perhaps, early stages of the development of expertise can be characterized by a familiarity with many components and procedures, whereas the mapping of

those components and procedures to a particular troubleshooting situation is a hallmark of later stages of expertise.

Finally, it is interesting that the laddering technique was predictive of troubleshooting performance only in the context of the troubleshooting problem. Lists of the general system components or interfacing components were not predictive. Gitomer (1984) also used a laddering structured interview in this domain, however he did not restrict the interview to a particular problem context. Instead, subjects were told to think of a specific LRU and to iteratively break this LRU down into its components. This lack of context may explain why Gitomer did not observe a difference between skilled and less-skilled airmen (as defined by supervisor ratings) in the laddering technique.

Relatedness Ratings

Technicians' familiarity ratings indicated that, on average, they were familiar with the RWR system components ($\underline{M}$ = 1.9, $\underline{SD}$ = .919, collapsing across components and technicians). However, two 3-level technicians each rated two components as unfamiliar (rating = 6). First, they both rated the LRU11 as unfamiliar. This component is a control panel with which technicians do not regularly interact. Perhaps these two technicians had not yet come into contact with the LRU11 in the course of troubleshooting. Furthermore, it is not an important component for troubleshooting the RWR Wiring--Broken Wire problem. The two technicians also rated different antennas as unfamiliar. One technician rated the low-band antenna as unfamiliar, and one rated the right fin antenna as unfamiliar. These ratings are unusual, particularly the rating given for the right fin antenna. The right fin antenna is one of the four high-band antennas. The technician rating the right fin antenna as unfamiliar rated the remaining three high-band antennas as familiar. These antennas perform the same function in different areas of the aircraft. Being familiar with one of the high-band antennas implies familiarity with the remaining three high-band antennas. Perhaps these ratings could be attributed to input error. Regardless, on average, the technicians were very familiar with the RWR system components.

Correlations of relatedness ratings for each pair of the four high performers were computed to determine degree of inter-subject agreement. These correlations are presented in Table 6. Note that the correlations are all high and statistically reliable ($\underline{p}$ < .05 with 54 degrees of freedom).

In order to generate a graphical summary of the ratings, the data were submitted to the Pathfinder network scaling procedure, a descriptive multivariate statistical technique that represents pairwise proximities in a network form (Schvaneveldt, 1990). In the networks, concepts (or components in this case) are represented as nodes, and relations (functional relations in this case) are represented as links between nodes. This network representation not only summarizes the data but has been shown to convey invormation about conceptual relatedness not

seen in the ratings themselves (Cooke, 1992; Cooke, Durso, & Schvaneveldt, 1986). For more detail on Pathfinder see Schvaneveldt (1990) or Cooke and Rowe (1993). The C statistic (Goldsmith & Davenport, 1990), a measure of shared links for matching nodes across two different networks, was calculated between each of the four high performers (Table 7). This measure ranges from 0 (low similarity) to 1 (high similarity) and can be viewed as a measure of association between two networks. The C values in Table 7 indicate that technician #8 shares fewer links with the other three high performers than the other three share with each other. Therefore the standard was computed in several ways: (1) averaging ratings of all four high performers, (2) using ratings of technician #8 only, and (3) averaging the ratings of all high performers, excluding #8. The resulting Pathfinder network based on the data from the second standard is presented in Figure 4a. For comparison purposes Figure 4b presents the Pathfinder network based on the mean relatedness ratings of the eight lowest performers (i.e., troubleshooting score less than 50).

Table 6

Intercorrelations of relatedness ratings for the four high performers.

	Technician Number		
	6	8	14
5	.757	.769	.889
6		.662	.751
8			.899

The knowledge index for each of the 15 technicians was based on the C value between the individual technician's network and each of the three standards described above. Correlations were then computed between knowledge indices and troubleshooting scores to determine which, if any, of these three knowledge indices predicted performance. Correlations were respectively, .399, .527, and .425 for the first, second, and third standards described above. Only the second correlation is significant ($\alpha_{TW} = .03$, df = 13), indicating that only the knowledge index using technician #8 as the standard (standard two) was predictive of performance. A partial correlation ($\underline{r}$ (12) = .405, $\underline{p}$ = .06), although only marginally significant, suggested a knowledge index based on technician #8 was predictive of performance even when the variance accounted for by the remaining three high performers (i.e., the third standard) was partialled out. On the other hand, the correlation between performance and the standard based on all four high performers (i.e., first standard) dropped from .399 to -.034 when the variance due to technician #8 was partialled out.

Table 7
C values for each pair of the four high performers' networks.

	Technician Number		
	6	8	14
5	.887	.418	.868
6		.423	.759
8			.404

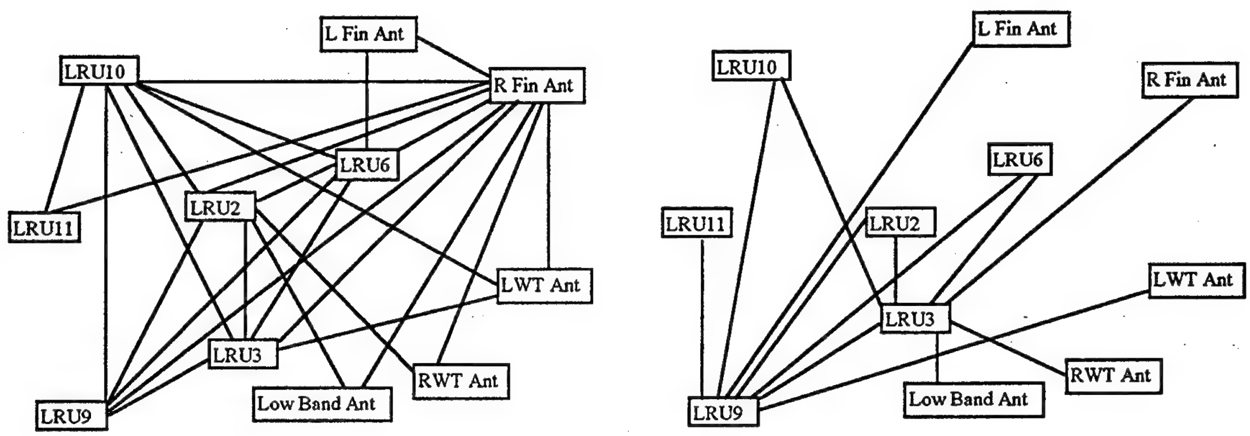

Figure 4a Figure 4b

Figure 4. Pathfinder networks (r = ∞, q = n-1) based on: (a) the data from technician #8 - the second standard and (b) the mean relatedness ratings given by the eight lowest performers (score < 50). Note: link weights have been omitted. Link length does not indicate link strength.

Thus, relatedness ratings, coupled with the Pathfinder network analysis procedure, were predictive of troubleshooting performance when technician #8's ratings were used as the standard. This technique, however, was only marginally predictive of troubleshooting performance when all high performers were used as the standard. The C values for the high performers, in combination with the marginally significant partial correlation, indicate that there is little overlap between the ratings given by technician #8 and those of the other three high performers. Interestingly, there are some other differences between technician #8 and the other high performers. That is, #8 is a 5-level, whereas the others are all 7-levels. Also, #8 has spent only four years in the C shop, whereas the others have spent between 6 and 8 years in the C shop. However, #8 performed as well as the other high performers in the troubleshooting task, receiving a score of 95 compared to

the other three scores of 95, 97, and 100. Possible explanations for this pattern of results are discussed in the General Discussion section.

Diagramming Task

The following analyses are based on the general diagrams of the RWR system. The diagrams that were specific to the RWR Wiring--Broken Wire problem were not informative because the majority of the technicians deemed that only three to five system components were relevant to this problem. Each of the 19 technician's system diagrams was converted to an 11 by 11 asymmetric matrix, with ones representing the presence of a connection between components and zeros indicating no connection. To determine the level of diagram similarity among the four high performers the proportion of shared connections for each pair of diagrams was computed. This proportion was based on the number of links shared by the two diagrams divided by the total number of links in the union of the two diagrams. These proportions are presented in Table 8. In general, pairs of high performers shared about half of the links present in the two diagrams. However, closer inspection of Table 8 indicates that once again, the diagram of technician #8 shared the least with those of the other high performers (mean proportion of .42 for technician #8 compared to .68 for other pairs of high performers). For this reason, three standard diagrams were created parallel to the three standards for relatedness ratings: (1) a diagram based on all four high performers, (2) technician #8's diagram, and (3) a diagram based on all high performers excluding #8. The matrix representing the group diagrams (i.e., standards one and three) consisted of ones, indicating that the connection existed in at least one diagram, and zeros otherwise. Again these matrices were asymmetric.

Table 8

Proportion of shared connections for pairs of high performers' diagrams.

	Technician Number		
	6	8	14
5	.55	.48	.85
6		.39	.65
8			.40

Three diagramming knowledge indices were generated for each of the remaining 15 technicians by subtracting the technician's matrices from each of the three standard matrices and summing the absolute values of the differences. These indices should be zero if there is complete agreement with the standards. The correlations between these knowledge indices and trouble-shooting performance were -.464, -.440, and .089 for standard one, two, and three, respectively.

The first two correlations are significant ($\underline{p}$ = .041 and .051, respectively, α_{TW} = .03) with 13 degrees of freedom. Note that a negative correlation is expected given that large knowledge indices indicate large diagram differences. The low correlation between standard three and performance, as well as a nonsignificant correlation of -.181 when technician # 8's diagram is partialled out of the first standard, suggests that similarity to technician #8 is responsible for the predictability of this technique. In fact, when the three other high performers are partialled out of the correlation between technician #8 and performance, the resulting correlation is high and significant ($\underline{r}$ (12) = -.538, $\underline{p}$ = .02). Thus, as for relatedness ratings, prediction of technician performance is optimal with a knowledge index based on technician #8. The diagram completed by technician # 8 is presented in Figure 5.

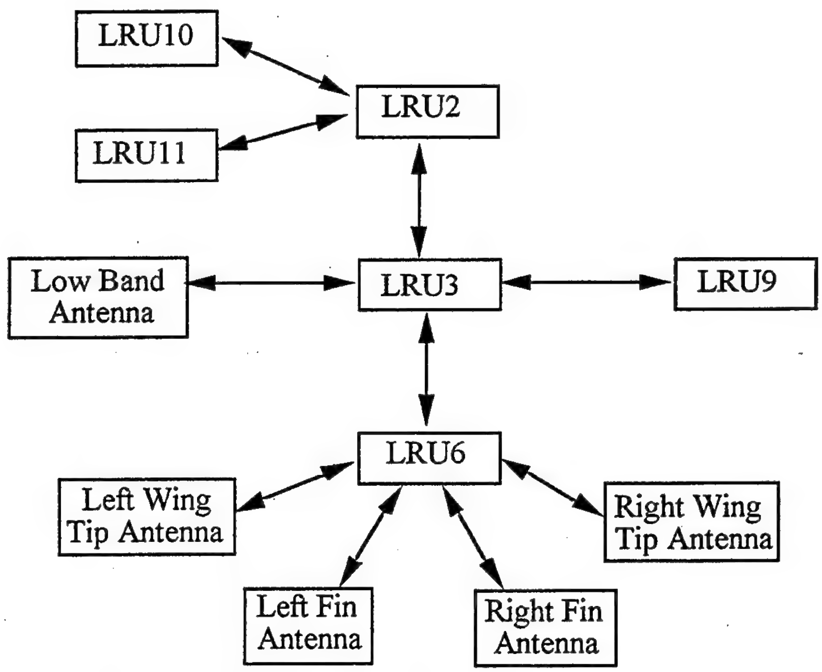

Figure 5. RWR system diagram created by technician # 8.

Interestingly, using #8 as the standard, errors of omission ($\underline{r}$ (13) = -.508, $\underline{p}$ =.026) are better predictors of performance than errors of commission ($\underline{r}$ (13) = -.272, $\underline{p}$ =.167). This result indicates that those technicians whose diagrams did not include connections seen in technician #8's diagram tended to exhibit poorer troubleshooting performance. It is interesting that errors of omission should be predictive as opposed to commission, given the laddering task finding in which errors of commission were more predictive of performance. However, these two results may not be disparate. Specifically, the fact that errors of ommission in the diagramming task were predictive suggests that seeing "extra" system relations is not as problematic for troubleshooting as is failing to see one or more critical relations. Similarly, the fact that errors of commission in

the laddering task were predictive suggests that the best troubleshooters believed that many components were relevant for troubleshooting the problem, including those which are actually relevant. Thus, better troubleshooters appear to have a wide range of information available, a subset of which is considered critical by experts. These results also indicate that the laddering and diagramming tasks may be accessing different aspects of mental model knowledge.

In general, the diagramming technique, at least in the context of the overall system, predicted troubleshooting performance well. Predictability was best when technician #8 was used as the standard, the same technician whose relatedness ratings predicted performance. In terms of the diagramming technique, technician #8 differed from the other high performers in two major ways: (1) #8 used only bi-directional arrows, and (2) based on technicians' diagram explanations, #8 attempted to represent both information flow and power flow in his diagram, whereas the other high performers represented only information flow in their diagrams.

Think Aloud

A coding scheme for technicians' verbalizations in the think aloud while troubleshooting technique was developed. The purpose of the coding scheme was to be able to classify verbalizations into discrete meaningful units that could be represented as nodes in a Pathfinder network. The main groups of verbalizations included: (1) action interpretation/explanation, (2) result interpretation, (3) component elimination, (4) elimination justification, (5) plan/prepare for test/check, and (6) Technical Order (T.O.) search/interpretation. The most abstract level of categorization was used for each category except for the action interpretation/explanation category. This category was broken down into sub-units because the type of action interpretation/explanation seemed important in distinguishing skill levels. For example, if a technician checked the fuses on the LRU2 (a power check) and verbalized an explanation other than a power explanation, this information would likely distinguish different skill levels and should be captured. The resulting coding scheme consisted of 22 verbalization units/nodes. Each technician's verbal protocol was coded by two raters. The inter-rater reliability achieved on the recoding of the action interpretation/explanation category was acceptable, with 92.6% agreement on 149 coded explanation/interpretation verbalizations. The raters discussed the 11 verbalizations on which they disagreed, and a compromise was made.

Also included in the coding scheme were meaningful troubleshooting actions that technicians executed. These were included in order to provide context for the verbalizations. Meaningful was defined as actions indicative of skill in troubleshooting the presented problem. For example, checking the LRU9 ETI meter (a power indicator) is indicative of troubleshooting skill for the RWR Wiring--Broken Wire problem because power to the LRU9 may be the cause of the problem. On the other hand, checking the LRU6 ETI meter is not because problems with the LRU6 are not indicated by the problem statement. The main groups of actions included: (1)

debriefing questions, (2) equipment checks, (3) continuity tests, and (4) swaps. The most abstract level of categorization that indicated troubleshooting skillfulness was used. Using this decision rule, an action unit was associated with either poor or good troubleshooting actions. The resulting coding scheme consisted of 75 action units/nodes. Two raters coded 5 of the action protocols together. They then coded the remaining 14 action protocols and achieved an acceptable level of reliability of 98.1% with 267 coded actions. The raters discussed the five actions on which they disagreed, and a compromise was met. The entire coding scheme (verbalizations and actions) included 97 events/nodes (22 verbalizations + 75 actions) and can be seen in Appendix E.

The coded protocols were then analyzed in terms of transition probabilities for pairs of utterances. In general, this analysis focuses on recurring sequential patterns in the coded protocols. Transition probabilities for all event pairs (verbalizations and actions) were calculated for individual subjects by dividing the frequency with which specific event transitions (e.g., T.O. search/interpretation followed by continuity check between LRU2 and LRU3) occurred by the frequency with which the first event in the sequence occurred. For example, if T.O. search/interpretation occurred twice and was followed by continuity check between LRU2 and LRU3 on one of those occasions, then the transition probability would be 0.5. Note that these are first-order transitions only. Higher-order transitions were not used because the immediate transitions were considered to be the most meaningful for this task. Also note that event transitions convey order information, and thus each event pair can be associated with two distinct transition probabilities.

Transition probabilities were calculated for each of the four high performers. Agreement among these technicians was assessed by correlating these probabilities for all pairs of high performers. These correlations are presented in Table 9. Note that all inter-technician correlations are low and are not statistically significant. However, because there is no single technician who seems different from the rest, the transition frequencies of all four technicians were combined as the standard. The standard transition matrix was then submitted to the Pathfinder network scaling algorithm, as were transition matrices of each of the individual technicians (Schvaneveldt, 1990). Similarity measures (i.e., C values as used for relatedness ratings) among pairs of networks for high performers corroborated the correlation results, with a mean inter-technician C of .06. This low agreement suggests that high performers conveyed very different thoughts in their verbalizations.

C values between the standard Pathfinder network and each individual technician served as a knowledge index. This index was not predictive of troubleshooting performance ($\underline{r}$ (13) = -.026). Note that this is not a matter of idiosyncrasies among the high performers in that standards based on any one of the four are not significantly predictive of performance (correlations of performance

Table 9
Correlations of transition probability matrices for pairs of high performers.

	Technician Number		
	6	8	14
5	.05	.15	.06
6		.06	.01
8			.03

and knowledge indices based on individuals are -.325, -.005, -.133, and -.139). The highest observed correlation (between technician #5 and performance, r = -.325) is in the opposite direction, indicating that the best troubleshooters verbalized in a manner least like that of technician #5.

An additional knowledge index was derived from a correlation of event frequencies (i.e., the frequency with which each discrete verbalization or action occurred) associated with an individual's protocol and event frequencies associated with the standard. Thus, this measure should be high to the extent that the technician exhibited the same verbalizations and performed the same actions as the high-performers, the same number of times. It should overlap with the Pathfinder network similarity measure in that they both take shared events into account. However, these two measures are divergent in that the Pathfinder measure includes information on event sequences, whereas the action frequency measure does not. Conversely, the frequency measure includes frequency of individual events, whereas Pathfinder does not. This measure (r (13) = -.394, p =.07) was slightly more predictive than the network similarity measure, though not significantly so. However, this correlation is negative, indicating that technicians with verbalization/action frequencies similar to that of the standard had lower troubleshooting scores.

In summary, the think-aloud technique resulted in low levels of agreement among the four high performers and was not predictive of troubleshooting performance. The marginally significant negative correlation between the frequency knowledge index and performance suggests that the better technicians say different things at different rates than high performers. Specifically, the four high performers tended to say less than the other technicians. However, of the 15 technicians, those who performed well tended to read the Technical Orders and thus generate more verbal statements. Together these results seem to indicate that the think-aloud technique does not assess knowledge that is critical for performance. Technicians may be unaware of much of the knowledge that underlies their troubleshooting performance and thus, when asked to think aloud, they verbalize thoughts that are independent of task performance.

Questionnaires

Each of the questions on the Likert-scale questionnaire was analyzed using a repeated measures ANOVA, with the four techniques (i.e., laddering, ratings, diagramming, and think aloud) making up the four levels of the independent variable. Two of the five questions resulted in no significant technique effect. These questions had to do with similarity to actual troubleshooting and usefulness as a measure of system knowledge. The other three questions resulted in significant technique effect. For the task difficulty question, there was a significant effect of technique (F (3, 54) = 3.5, p = .02), that can be attributed to the fact that the laddering technique was rated as significantly easier than each of the other three techniques. Respective means for laddering, ratings, diagramming, and think aloud were 5.3, 4.2, 4.3, and 4.2. For the question about restriction in range of responses, there was also a significant technique effect (F (3, 54) = 6.5, p = .003). The ratings technique was viewed as significantly more restrictive than laddering and think aloud techniques. Respective means for laddering, ratings, diagramming, and think aloud were 4.5, 3.6, 4.4, and 5.0. For the task artificiality question there was a significant technique effect (F (3, 54) = 3.42, p = .02), with the think aloud and laddering techniques receiving greater realism ratings than the other two tasks. Respective means for laddering, ratings, diagramming, and think aloud were 4.0, 3.4, 3.4, and 4.6. Also, the 19 technicians, on average, viewed system knowledge as extremely important (mean rating of 4.3 on a 5-point scale, SD = .65) for troubleshooting the RWR Wiring--Broken Wire problem.

On the second questionnaire technicians circled the technique in each pair of techniques which they felt best measured the knowledge necessary for troubleshooting. Percent responses indicated that technicians viewed the diagramming (39%) and think aloud (33%) techniques as better in this regard than the laddering (12%) and rating (16%) techniques.

Supervisor ratings were to have been collected for each of the 19 technicians. However, ratings for technicians from one of the participating fighter squadrons were unavailable. Six technicians are members of this squadron. Thus, ratings were collected for only 13 of the technicians. Due to the missing data, supervisor ratings were not analyzed.

GENERAL DISCUSSION

Summary and Recommendations for Measuring Mental Models

The purpose of this work was to identify one or more techniques suitable for measuring mental models of the type relevant to avionics troubleshooting. Air Force technicians with varying degrees of expertise completed four mental model measures. These measures were conducted in the context of a specific troubleshooting problem. Care was taken in the selection of this problem to insure that it was representative of the domain (i.e., C shop avionics), and of intermediate difficulty, presumably making the invocation of a system mental model necessary for

successful troubleshooting. In addition to completing the mental model measures, each technician worked to verbally troubleshoot the problem. Comparing the results of the knowledge measures to troubleshooting performance should provide a pragmatic means of assessing the validity of the measures. Of the four techniques tested, all but the think-aloud technique were predictive of troubleshooting performance. Whether the think-aloud standard was constructed from individual high performers or groupings of these high performers, predictability remained low. On the other hand, the laddering (Step 2), relatedness ratings, and diagramming techniques were all predictive of troubleshooting performance. (Table 10 lists the correlations between these techniques and troubleshooting performance). Also, when technician #8 was used as the standard in the ratings and diagramming techniques, predictability was optimal. An explanation for this is discussed below.

Table 10
Correlations of troubleshooting performance (TS Score), laddering technique (Step2), ratings technique, and diagramming technique.

	Measurement Technique		
	Laddering	Ratings	Diagram
TS Score	.542	.527	-.440
Laddering		.288	-.477
Ratings			-.274

Partial correlations indicated that two of the techniques were each independently predictive of troubleshooting performance. Specifically, the relatedness ratings technique was predictive of performance independent of both the laddering technique (r (12) = .461, p = .042) and the diagramming technique (r (12) = .471, p = .039). Similarly, the laddering technique was predictive of performance independent of the ratings technique (r (12) = .480, p = .035) and to a lesser extent, the diagramming technique (r (12) = .421, p = .057). Interestingly, the diagramming technique was not predictive independent of the relatedness ratings (r (12) = -.362, p = .09) or the laddering (r (12) = -.246, p = .18) techniques. Thus, the ratings technique and the laddering technique each assessed mental model information that is independently predictive of performance, whereas the diagramming technique did not.

Based on the results of this research, several recommendations can be made regarding system knowledge measurement techniques. First, the relatedness ratings and laddering techniques each independently predicted troubleshooting performance, indicating that these two methods are valid means of assessing mental model knowledge. On the other hand, neither the diagramming technique or the think-aloud technique was independently predictive of

performance. Thus, researchers should consider using the relatedness ratings and laddering techniques in work requiring the measurement of mental models. Interestingly, this recommendation is counter to the technicians' ratings which indicated that they believed the diagramming and think-aloud techniques were the best measures of mental model or system knowledge.

Further recommendations can be made based on the subjective ratings given by technicians. Specifically, the laddering technique should be used in cases in which both ratings and laddering cannot be used. Technicians thought that the rating technique was the most restrictive in terms of response freedom and that it lacked realism. Procedurally, relatedness ratings are also restrictive in the sense that presenting all pairs of a concept set quickly leads to an unmanageable number of pairs as the number of concepts increases, making the ratings technique nearly impossible to use with large concept sets. Even with a smaller number of concepts (around 20), the pairwise ratings task seems quite long and tedious to subjects. On the other hand, the laddering technique appears to be easier to implement. The laddering technique requires less background knowledge on the part of the researcher than does the ratings technique, especially when ratings are followed by Pathfinder analyses. In addition, technicians rated the laddering technique as realistic and easy to complete. However, there is a tradeoff to be considered. The ratings technique, while relatively more difficult to implement and analyze than the laddering technique, provides a graphical representation of knowledge that is much richer than the list produced using the laddering technique. Furthermore, the problems associated with the ratings technique are less important when the ultimate goal of our research program is considered: the on-line prediction of knowledge via action patterns. That is, once a correspondence between distinct representations of system knowledge and distinct action patterns has been identified, the knowledge elicitation step may be bypassed.

The results obtained from this work are particularly promising when aspects of the study are considered. First, the sample size was minimal. The data were gathered in a "real-world" setting, increasing the ecological validity of the conclusions. However, subjects are not typically available in large numbers in such settings, particularly when they must leave their daily work to participate. Only nineteen subjects participated in this study. Furthermore, data from four of these subjects were extracted for use in construction of the standard, restricting the range of data on which the correlations were based. Thus, the correlations reported here may be underestimated. Finally, the purpose of this work must be emphasized: the interest was an evaluation of the validity of several mental model measures, with the criterion being troubleshooting performance. The observed correlations between mental model measure performance and troubleshooting performance varied across the evaluated techniques, indicating that not all mental model measurement techniques are created equal. Some measures are predictive of performance, whereas others are not. Measures

predictive of performance should be pursued in future work. In other words, given a selection of mental model measures, it would pay to use those measures that produce output that seems most relevant to performance.

What remains to be done is an in-depth examination of the information provided by the various methods for measuring mental models. A characterization of the types of information provided by the different methods for measuring mental model knowledge would assist researchers in choosing the most appropriate method or methods to meet their research needs. For example, the laddering interview task appears to access knowledge of system components, whereas the ratings task appears to measure knowledge of relationships among system components. Researchers may find these types of information differentially useful, depending on their research needs. Conducting an in-depth analysis of the types of information provided by different mental model measures would allow researchers to select the measure(s) best suited to their research needs. Furthermore, such a characterization would serve to strengthen the network of evidence (or nomological network, Cronbach & Meehl, 1955) surrounding the mental model construct. For example, different measures, though equally predictive of performance, may tap different aspects of mental model knowledge. An understanding of the types of information provided by the different measures would facilitate the integration of research results, increasing understanding of the mental model construct in general.

Findings on Measuring Mental Models

Defining the Standard. This research revealed several interesting observations relevant to mental model measurement and expertise in general. The first, and perhaps most important, observation dealt with defining the "standard." We found that using a single technician, #8, as the standard resulted in optimal predictability in the ratings and diagramming techniques. As mentioned earlier, technician #8 differed from the other high performers on two dimensions: years of experience and level obtained in the Air Force classification system. This technician may be a very good "intermediate" level technician rather than an "expert" level technician. Perhaps using technician #8 as the standard resulted in better predictability because the other high performers were too far removed from the entry-level technicians in terms of expertise, whereas #8 was not. Thus, although technician #8 and the other three high performers each performed well on the troubleshooting problem, their performance on the knowledge measures was different. Perhaps technician #8's knowledge was most like that of a good intermediate-level technician. Using experts as the standard assumes a linear relationship between the development of expert performance levels and the development of knowledge. Perhaps major qualitative changes in knowledge occur as expertise develops, making expert knowledge a poor predictor of novice performance.

Such qualitative changes in knowledge are central to the perspective taken by phase learning theorists (Shuell, 1990). Phase theorists assert that learning a complex body of knowledge involves a series of phases, during which the learning process is fundamentally different. Furthermore, although it is typically assumed that the phases are organized in a linear manner, a non-linear organization is possible. Thus, changes in knowledge may not be adequately represented as a simple monotonic increase in similarity to some ideal knowledge representation (Acton, Johnson, & Goldsmith, 1994). Instead, the knowledge organization of a subject at one expertise level may be qualitatively different from that of another subject at a different expertise level, making unique "standards" or ideals necessary for each phase of knowledge development.

An analysis conducted on unpublished data provided by Cooke (1994) offers support for this point. Cooke collected pairwise similarity ratings on a set of cognitive psychology concepts from undergraduate students enrolled in a cognitive psychology class. Two sets of ratings were collected: one set at the beginning of the semester and one set at the close of the semester. The course instructor also completed the ratings. In order to assess the predictive power of different standards or ideals, two standards were evaluated: (1) the ratings given by the instructor and (2) an aggregate high-performer rating set created by averaging the second set of ratings given by the top 5 students in the class (as determined by final grades). Separate analyses were conducted on the first and second set of ratings given by the remaining students. The second set of ratings given by individual students were correlated with both the instructor's ratings and the average high-performer ratings given at the end of the semester. These values were then correlated with final grades to determine if the standards differed in their respective abilities to predict performance. The resulting correlations indicated that both standards were predictive of class performance, $\underline{r}$ (62) = .486, $\underline{p} < .0001$, and $\underline{r} = .645$, $\underline{p} < .0001$ for the instructor and the high performer standards, respectively.

Partial correlations were then conducted to determine if the standards were independently predictive of performance. The resulting partial correlations indicated that the high-performer standard was predictive of performance independent of the instructor standard, $\underline{r}$ (61) = .486, $\underline{p} < .01$. However, the instructor standard was not predictive of performance independent of the high-performer standard, $\underline{r}$ (61) = .030. Thus, students who gave ratings similar to those given by the high performers were more likely to succeed in the class. Giving ratings similar to those of the instructor was not independently related to class performance. Interestingly, a comparable analysis conducted on the students' first set of ratings indicated a similar trend. Specifically, students whose first set of ratings were like the second set of ratings given by the high performers tended to perform well in the class, $\underline{r}$ (62) = .266, $\underline{p} = .04$. On the other hand, there was no relationship between having ratings similar to those given by the instructor and class performance,

$r(62) = .152$, $p = .24$. Taken together, these results indicate that an ideal based upon high-performing students offers more predictive power than an ideal based upon the expert instructor.

Similar results were obtained in a study conducted by Acton, Johnson, and Goldsmith (1994) who examined different structural knowledge referents or standards in terms of their abilities to predict exam performance. These researchers collected pairwise similarity ratings on a set of programming concepts from students enrolled in one of three BASIC programming classes. Ratings were also collected from the course instructors and from a group of non-instructor experts. Four standard structures were evaluated: (1) the course instructor's ratings, (2) individual experts' ratings, (3) an average of the ratings given by the non-instructor experts, and (4) averages of the ratings given by the six students receiving the highest marks in each of two programming classes.

The results indicated, among other things, that the standard based on the six best students in the class was a slightly better predictor of exam scores for that class ($r(26) = .35$, $p < .10$ and $r(31) = .57$, $p < .05$ for Classes 1 and 2, respectively) than the standard based upon the course instructor ($r(26) = .33$, $p < .10$ and $r(57) = .55$, $p < .05$ for Classes 1 and 2, respectively). Furthermore, when the standards based upon the students in Classes 1 and 2 were used to predict performance in a third, more advanced class, the resulting correlations were larger ($r(8) = .46$ and $r(8) = .45$, respectively) than when the course instructor was used to predict performance ($r(8) = .35$), although none of these correlations are significant.

Furthermore, an examination of the correlations resulting from the standards based on the non-course instructor experts indicates a similar trend for two of the three classes. That is, the student-based standard predicted in a manner similar to or slightly higher than the standards based on the non-instructor experts. However, in one class the non-instructor expert standards were more predictive of performance. In general, these results indicate that standards based upon high-performing students offer comparable or superior predictive power relative to standards based upon experts.

Taken together, the results of these studies indicate that the knowledge organizations of students are qualitatively different from the knowledge organizations of experts. Experts appear to be at a qualitatively different stage or phase of learning and understanding, one which is not predictive of student performance. Expert knowledge structures may be too far removed from student knowledge organizations to allow for great predictive power. High-performing or high-level intermediate students, however, may be at a more advanced position in the same stage of learning, allowing more powerful predictions. Thus, it appears that using a unique ideal for different stages of knowledge development improves predictive power.

Although a standard based on technician #8 resulted predictability for the ratings and diagramming techniques, such a standard was not predictive for two of the evaluated mental

model measurement techniques. First, when technician #8 was used as the standard for the think-aloud technique, the technique was not predictive of performance. This lack of predictability may be more indicative of the think-aloud technique rather than using technician #8 as a standard. Verbalizing thoughts appeared to be problematic for the technicians. For example, some of the more experienced technicians did not verbalize, while other technicians simply read the T.O. out loud. In general, information that was verbalized was highly variable, and little of it seemed directly related to task performance. Second, a laddering interview knowledge index using technician #8 as a standard was not predictive of troubleshooting performance. Here, technician #8 did not perform differently from the other high performers. Perhaps the laddering technique taps into very basic system knowledge about existing components which may not evolve with expertise to the same extent as does knowledge about component interrelations which is tapped by the ratings and diagramming techniques.

Providing a Context. A second interesting observation deals with the context provided during mental model measurement. Specifically, providing a troubleshooting problem context when measuring mental model knowledge produced conflicting results. For the laddering technique, providing a problem context was useful. Predicting troubleshooting performance was greatest when components relevant to the troubleshooting problem were listed (Step 2). However, listing all system components regardless of problem context (Step 3) was not predictive of troubleshooting performance. On the other hand, providing a problem context in the diagramming technique resulted in diagrams which were less predictive of performance than were diagrams constructed without restriction to the specific troubleshooting problem. This diagramming finding could be due to the nature of the troubleshooting problem. Specifically, there were very few diagram components that were relevant to the problem. It may also be that task order is responsible for these different problem-context effects. Context was helpful in the laddering technique where the context-specific questions preceded the general ones and not in the diagramming technique where the general questions came first. Finally, the differential context effects observed for these two measures suggests that the measures may be accessing different aspects of mental model knowledge, one which is sensitive to context and one which is not. Once again, a characterization of the types of information provided by the various mental model measures would be useful. Here, it may provide some insight into the reasons for these context effects.

Development of Expertise. Finally, this research appears to support a pattern of development in troubleshooting expertise observed by Cooke and Rowe (1993) who examined the actions taken by technicians tasked with troubleshooting complex avionics equipment. They found that, after training, low performers exhibited a wide range of troubleshooting actions. A portion of these actions were executed by high performers, and a portion were not. These results

suggest that prior to achieving expertise, but after some experience, technicians have an extensive array of knowledge in the form of executable actions, but they do not know when these actions apply. Results from the current study support this conclusion and suggest that in fact this pattern is necessary in the development of expertise. The best technicians shared laddering components with the standard, but they also made more errors of commission. The fact that commission errors are predictive of performance indicates that the acquisition of this "extra" knowledge may, in fact, be a necessary stage in the development of expertise. Also, in the diagramming technique technicians with fewer omissions relative to the standard were more likely to be better troubleshooters. Again, the better troubleshooters may have a wide range of information, a subset of which is information considered critical by experts.

This phenomenon is also illustrated by a comparison of the network generated from technician #8's ratings and the aggregate network created from the eight lowest performers' ratings (see Figures 4a and 4b). First, the low performers focused exclusively on two system components: the LRU9 and the LRU3. Notice that all connections involve one or both of these components. These two components are also central in the network of technician #8, however, other appropriate components (e.g., the LRU2) are also central. Technician #8's network also includes many more links (including many weak ones, link weights > 3.0) than the low performers' network. Thus, with experience, it appears that technicians not only learn more executable actions (Cooke & Rowe, 1993), but technicians also appear to learn more interrelations among system components.

Interestingly, these results correspond with Karmiloff-Smith's (1986) observations of children's cognitive and linguistic development. For example, in her studies of children's acquisition of French grammar she noted that error rates change with increasing knowledge. Specifically, error rates are low during early acquisition (around five-years old). This correct usage is followed by a period (between approximately five- and seven-years old) when error rates increase; grammatical output becomes markedly different from that of adult grammar. Correct usage is observed once again at approximate eight-years old. Lesgold et al. (1988) noted that radiologists varying in expertise exhibit similar error patterns. That is, on some of the film stimuli more advanced radiologists were less likely to offer correct interpretations than were less advanced radiologists. Together, these results indicate that performance may not be a monotonic function of experience.

Karmiloff-Smith (1986) developed a three-phase model of the processes underlying linguistic and cognitive development based on her observations of children. During the first phase, learners are predominantly data-driven; they are working to create a match between the evaluation of their outputs (e.g., grammar) and their representations of the ideal (adult grammar). Few errors are observed. During the second phase, learners are actively working to alter and

organize their internal representations. Because the learner's focus is on internal representations rather than the external environment, errors tend to increase. When the third phase is reached, learners have well-developed internal representations. Error rates similar to those occurring during the first phase are observed, but the representations underlying responses are richer and more organized. Shuell (1990) developed a similar three-phase model, describing the processes involved in meaningful learning. Incorporating such a phase approach may enhance our understanding of the development of avionics troubleshooting expertise and the development of expertise in general.

Future Directions

In this project two effective methods of measuring system knowledge were identified. Each of these measures produced results that corresponded to troubleshooting performance. In previous work (Cooke & Rowe, 1993) meaningful action patterns that corresponded to troubleshooting performance were also identified. Also, note that Cooke and Rowe's results were replicated here with data collected for the RWR Wiring--Broken Wire troubleshooting problem. Event transition probabilities based on the coding scheme for actions only (see Think Aloud results) were generated and submitted to Pathfinder analyses. An action network from a subset of three high performers who agreed with each other was predictive of technician performance (r (13) = .415, p = .06). Thus, technicians' troubleshooting performance can be reliably predicted from relatedness rating and laddering measures as well as from action sequences.

Although the research completed thus far is valuable on its own (i.e., in terms of the identification of several techniques that can be used to predict performance off-line), the ultimate goal of our research program is a method for predicting knowledge during on-line tutor interactions. Thus, the next step toward this goal is to map action patterns onto knowledge measures. In other words, can patterns of actions be identified that correspond with specific mental models of the system? This mapping of actions to knowledge will enable technicians' mental models or deficits in system knowledge to be identified without requiring technicians to complete the laddering or ratings tasks. Instead, by relating action patterns to knowledge patterns, knowledge can then be predicted based only on action sequences.

Future work will involve collecting knowledge and action measures on the same group of subjects and identifying categories of action and knowledge patterns (i.e., types of mental models and associated actions). System or mental model knowledge can be predicted on-line to the extent that action categories map onto knowledge categories. Some analyses carried out on the data collected for this study are promising in this regard. Cluster analyses were performed on measures of inter-subject similarity for the action sequences as well as for the laddering (Step 2), ratings, and diagramming techniques. The action data resulted in one tight cluster of six subjects. If actions predict knowledge, then the technicians in this cluster who performed similar actions

should also be clustered together in analyses of knowledge measures. Indeed, three of these six technicians were in the same cluster in the laddering (Step 2) and diagramming analyses, and five of the six were together in the ratings analysis. More work on matching actions with knowledge should not only reveal the extent of the match, but also the specific nature of the knowledge categories (or mental models) and action patterns that characterize the population of technicians. These issues and others remain for future research.

REFERENCES

Acton, W.H., Johnson, P.J., & Goldsmith, T.E. (1994). Structural knowledge assessment: Comparison of referent structures. Journal of Educational Psychology, 86, 303-311.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. Science, 228, 456-462.

Burton, R. R. (1982). Diagnosing bugs in a simple procedural skill. In D. Sleeman & J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 157-183). London: Academic Press.

Cooke, N.J. (1992). Predicitng judgment time from measures of psychological proximity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 640 - 653.

Cooke, N.J. (in press). Varieties of Knowledge Elicitation Techniques. International Journal of Human-Computer Studies.

Cooke, N.J. (1994). [Predicting performance and misconceptions in cognitive psychology]. Unpublished raw data.

Cooke, N.J., Durso, F.T., & Schvaneveldt, R.W. (1986). Recall and measures of memory organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 538 - 549.

Cooke, N.J., & Rowe, A.L. (1993). An Approach to On-line Assessment and Diagnosis of Student Troubleshooting Knowledge. Air Force Technical Report, AL-TP-1993-0004. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

Gitomer, D.H. (1984). A cognitive analysis of a complex troubleshooting task. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Gitomer, D.H. (1992). Foundations of HYDRIVE. Presentation given at the Basic Job Skills Progress Review, Brooks AFB, San Antonio, April 7-9.

Glaser, R., Lesgold, A., Lajoie, S., Eastman, R., Greenberg, L., Logan, D., Magone, M., Weiner, A., Wolf, R., & Yengo, L. (1985). Cognitive task analysis to enhance technical skills training and assessment. (Final Report to the Air Force Human Resources Laboratory on Contract No. F41689-8v3-C-0029.) Pittsburgh: Learning Research and Development Center, University of Pittsburgh.

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. Schvaneveldt (Ed.), Pathfinder Associative Networks: Studies in Knowledge Organization (pp. 75-87). Norwood, NJ: Ablex.

Gott, S. P. (1989). Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In E. Z. Rothkopf (Ed.), Review of Research in Education, Vol 15, (pp. 97-169). Washington D.C.: American Educational Research Association.

Gouley, J.R.(1992) Integrated Avionics Communications, Navigation, and Penetration Aids System Specialist (F-15). Gunter Air Force Station, AL: Extension Course Institute Air University.

Gray, S.H. (1990). Using protocol analyses and drawings to study mental model construction during hypertext navigation. International Journal of Human-Computer Interaction, 2, 359-377.

Hall, E.M., Gott, S.P., & Pokorny, R.A. (1994). A Procedural Guide to Cognitive Task Analysis: The PARI Methodology. Manuscript submitted for publication.

Hall, E.M., Pokorny, R.A., & Kane, R. (1994, in preparation). Identification of Instructional Goals: A Qualitative and Quantitative Analysis of Problem Solving Protocols. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Karmiloff-Smith, A. (1986). Stage/structure versus phase/process in modelling linguistic and cognitive development. In I. Levin (Ed.), Stage and structure (pp. 164-190). Norwood NJ: Ablex.

Kieras, D.E. (1988). What mental model should be taught: Choosing instructional content for complex engineered systems. In J. Psotka, L.D. Massey, & S.A. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 85-118). Hillsdale, NJ: Erlbaum.

Larzelere, R.E., & Mulaik, S.A. (1977). Single sample tests for many correlations. Psychological Bulletin, 84, 557-569.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), The Nature of Expertise (pp. 311-342). Hillsdale, NJ: Lawrence Erlbaum Assoc.

McCloskey, M. (1983) Naive theories of motion. In D. Gentner & A.L. Stevens (Eds.), Mental Models (pp. 299-323). Hillsdale, NJ: Erlbaum.

Nichols, P.D., Pokorny, R., Jones, G., Gott, S.P., & Alley, W.E. (1989). Effective Instructional Processes Within an Intelligent Tutoring System. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division

Pokorny, R., & Gott, S. (1994, in preparation). The Evaluation of a Real-World Instructional System: Using Technical Experts as Raters. Air Force Technical Report. Brooks AFB, TX: Air Force Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Division.

Rasmussen, J. & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronics troubleshooting. Ergonomics, 17, 293-307.

Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.

Schvaneveldt, R. (Ed.), (1990). Pathfinder Associative Networks: Studies in Knowledge Organization. Norwood, NJ: Ablex.

Shuell, T.J. (1990). Phases of meaningful learning. Review of Educational Research, 60, 531-547.

Stevens, A., Collins, A., & Goldin, S.E. (1979). Misconceptions in student's understanding. International Journal of Man-Machine Studies, 11, 145-156.

Appendix A: The RWR Wiring--Broken Wire Problem

The Radar Warning Receiver (RWR) system is the system important in troubleshooting the RWR Wiring--Broken Wire problem. The RWR system consists of eleven components (see Table A1; Gouley, 1992). A functional block diagram of the RWR system components is illustrated in Figure A1. In this section, the general functions of each of the RWR system components are discussed. First, the LRU2 (Line Replaceable Unit) provides the RWR system with its power requirements. Specifically, the LRU2 receives 115 VAC 3 Phase 400 HZ and +28 VDC power from the aircraft and converts it to the DC voltages needed by the RWR. In addition, the LRU2 stores the flight program required for normal operation of the RWR. Next, the LRU3, in conjunction with the low-band antenna, provides the aircraft with coverage in the low-band frequency range. In addition, the LRU3 processes high- and low-band receptions, distributes data between components of the RWR, and interfaces the RWR with other avionics systems. The LRU6, in conjunction with the four high-band antennas (left wing tip, right wing tip, left fin, and right fin), provides the aircraft with omni-directional coverage in the high-band frequency range. The LRU9 provides all displays for the TEWS. It is located in the cockpit and provides the pilot with a view of the threat environment relative to the aircraft. The LRU10 is a control panel which allows the pilot to select on or off for the following systems: RWR, Internal Countermeasure Set (ICS), and Electronic Warfare Warning System (EWWS). The LRU11 is another control panel which allows the pilot to select RWR/ICS combat/training mode, mode of the TEWS pod, and ICS mode of operation.

Table A1.
RWR System Components

Component Name
LRU 2 (Power Supply)
LRU 3 (Low Band Receiver Processor)
LRU 6 (High Band Receiver)
LRU 9 (TEWS Display)
LRU 10 (TEWS control panel)
LRU 11 (TEWS immediate action control panel)
Left Fin Antenna
Right Fin Antenna
Left Wing Tip Antenna
Right Wing Tip Antenna
Low-Band Antenna

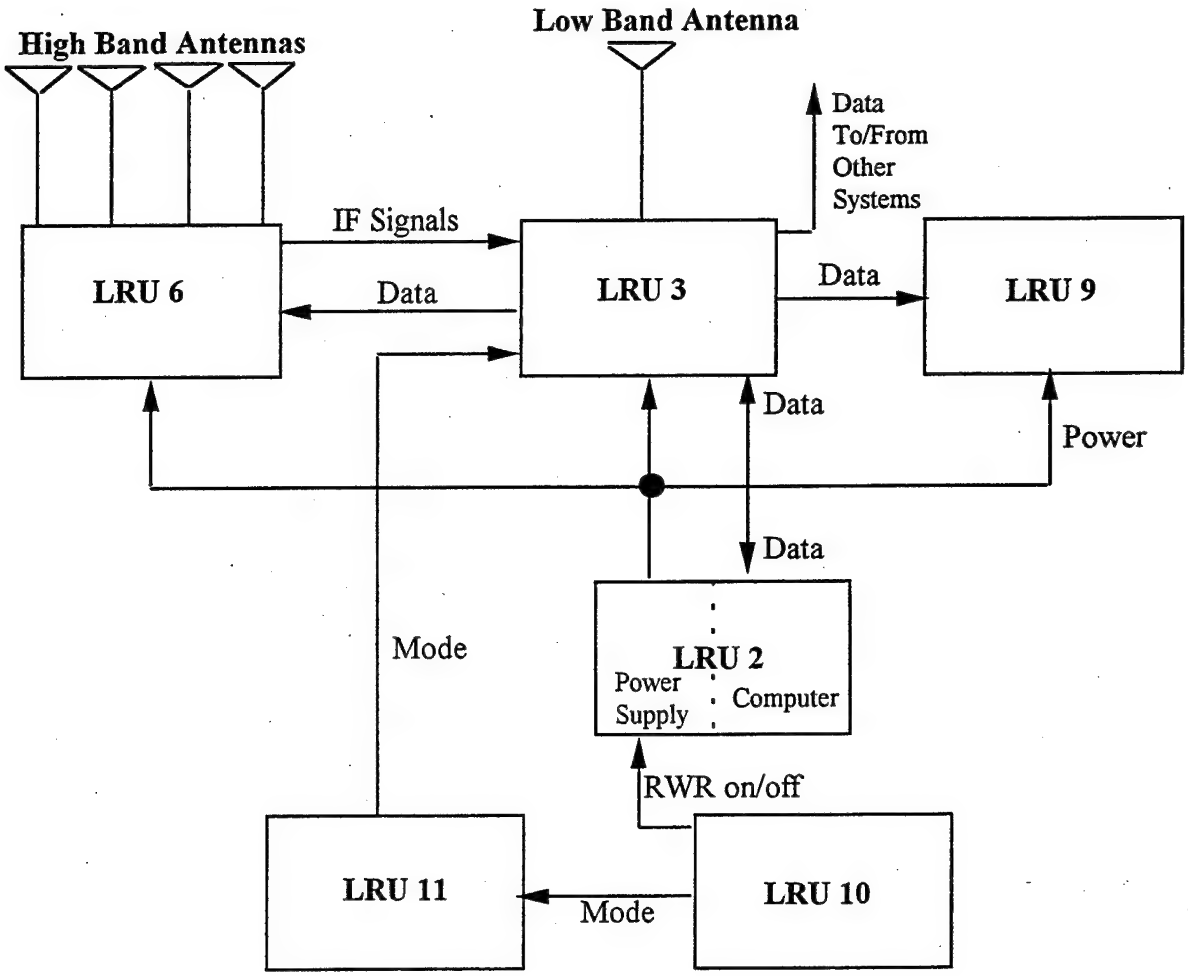

Figure A1. .Functional block diagram of the RWR system.

The RWR Wiring--Broken Wire problem is caused by a shorted video cable between the LRU3 and the LRU9. Specifically, the video cable between LRU3-2J8 and LRU9-9J1 is shorted. Center pin 1 on this cable is shorted to (touching) the shield on the LRU3 end, resulting in a loss of data (see Figure A2). When functioning correctly, this cable transmits video data from the LRU3 to the LRU9. The transmitted data contains information about the type and placement of symbols on the LRU9's display. However, due to the short between pin 1 and the shield, the data to be transmitted to the LRU9 are lost, and the LRU9 is completely blank.

Based on the problem statement (Table A2), technicians troubleshooting this problem may examine the following areas as possible causes of the problem: LRU9, LRU3, LRU2, or aircraft wiring. Without properly diagnosing the symptoms, the technician may leave the active signal path. For instance, the technician may perceive that the LRU9 is blank because there is no power to the system. Such a perception may lead to an investigation of the circuit breaker panel, TEWS

control panel, or the LRU2. A correct diagnosis, in addition to knowledge of the visual cues, can lead the technician to fault code 9911H1CA (available to technicians as part of the Technical Order or T.O.). This fault code lists the procedure to be followed given the following symptoms: no cross symbol or azimuth dots on TEWS display when power is applied to the RWR. The fault isolation (F.I.) tree for this fault code instructs the technician to check the data cable between the LRU3 and LRU9. Here, it is important that the technician has knowledge about the correct testing procedure of this cable. The F.I. only specifies to test for continuity across the center conductors from end-to-end for this cable. However, in order to detect the fault, the technician must thoroughly test the cable, which would include checking for continuity from center conductor to shield. Such a test would reveal that there is continuity here, indicating a short.

Table A2.
Problem Statement for the RWR Wiring--Broken Wire Troubleshooting Problem.

"In debrief, the pilot reports that the RWR is inoperative, the BIT (built-in test) light is on, and the TEWS display is blank."

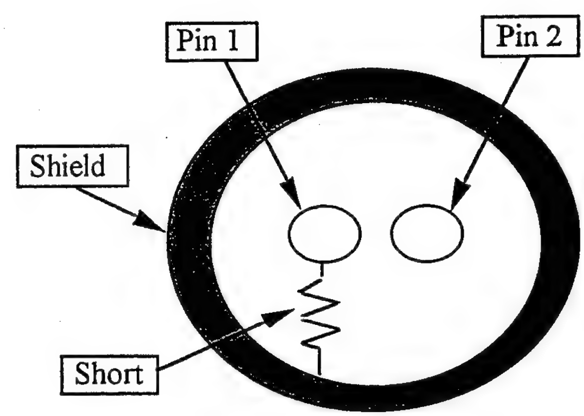

Figure A2. Cross-sectional view of the LRU3 end of the video cable between LRU3-2J8 and LRU9-9J1, in the context of the RWR Wiring--Broken Wire problem. Center pin 1 is shorted to the shield.

Appendix B: Questionnaire 1

Component Naming Task

(The following questions pertain to the task in which you generated the troubleshooting problem sytem components. These components were written on index cards and laid out in front of you as you named them.)

1. This task was: difficult--1--------2--------3--------4--------5--------6--easy

   Please describe why you rated the task as you did :______________________________

   ____

   ____

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1--------2--------3--------4--------5--------6--similar

   Please describe why you rated the task as you did :______________________________

   ____

   ____

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1--------2--------3--------4--------5--------6--broad

   Please describe why you rated the task as you did :______________________________

   ____

   ____

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1--------2--------3--------4--------5--------6--realistic

   Please describe why you rated the task as you did :______________________________

   ____

   ____

5. The information gained from this task is______________for measuring your System Knowledge.

   useless--1--------2--------3--------4--------5--------6--useful

   Please describe why you rated the task as you did :______________________________

   ____

   ____

6. Please write your comments here and on the back of this page:

Ratings Task

(The following questions pertain to the task in which you rated the relatedness of pairs of system components, using the computer.)

1. This task was: difficult--1--------2--------3--------4--------5--------6--easy

   Please describe why you rated the task as you did :____________________________

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1--------2--------3--------4--------5--------6--similar

   Please describe why you rated the task as you did :____________________________

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1--------2--------3--------4--------5--------6--broad

   Please describe why you rated the task as you did :____________________________

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1--------2--------3--------4--------5--------6--realistic

   Please describe why you rated the task as you did :____________________________

5. The information gained from this task is______________for measuring your System Knowledge.

   useless--1--------2--------3--------4--------5--------6--useful

   Please describe why you rated the task as you did :____________________________

6. Please write your comments here and on the back of this page:

Diagramming Task

(The following questions pertain to the task.in which you arranged and connected a set of system components, using index cards and paper arrows.)

1. This task was: difficult--1--------2--------3--------4--------5--------6--easy

   Please describe why you rated the task as you did :______________________________

   ____

   ____

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1--------2--------3--------4--------5--------6--similar

   Please describe why you rated the task as you did :______________________________

   ____

   ____

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1--------2--------3--------4--------5--------6--broad

   Please describe why you rated the task as you did :______________________________

   ____

   ____

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1--------2--------3--------4--------5--------6--realistic

   Please describe why you rated the task as you did :______________________________

   ____

   ____

5. The information gained from this task is______________for measuring your System Knowledge.

   useless--1--------2--------3--------4--------5--------6--useful

   Please describe why you rated the task as you did :______________________________

   ____

   ____

6. Please write your comments here and on the back of this page:

Thinking Aloud Task

(The following questions pertain to the talking aloud we asked you to do as you were troubleshooting the problem.)

1. This task was: difficult--1--------2--------3--------4--------5--------6--easy

   Please describe why you rated the task as you did :______________________________
   ____
   ____

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1--------2--------3--------4--------5--------6--similar

   Please describe why you rated the task as you did :______________________________
   ____
   ____

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1--------2--------3--------4--------5--------6--broad

   Please describe why you rated the task as you did :______________________________
   ____
   ____

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1--------2--------3--------4--------5--------6--realistic

   Please describe why you rated the task as you did :______________________________
   ____
   ____

5. The information gained from this task is______________for measuring your System Knowledge.

   useless--1--------2--------3--------4--------5--------6--useful

   Please describe why you rated the task as you did :______________________________
   ____
   ____

6. Please write your comments here and on the back of this page:

Troubleshooting Task

(The following questions pertain to the task in which we askedyou to troubleshoot the RWR Wiring problem verbally, with Sgt. Kruse reporting the equipment states to you.)

1. This task was: difficult--1--------2--------3--------4--------5--------6--easy

   Please describe why you rated the task as you did :______________________________
   ____
   ____

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1--------2--------3--------4--------5--------6--similar

   Please describe why you rated the task as you did :______________________________
   ____
   ____

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1--------2--------3--------4--------5--------6--broad

   Please describe why you rated the task as you did :______________________________
   ____
   ____

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1--------2--------3--------4--------5--------6--realistic

   Please describe why you rated the task as you did :______________________________
   ____
   ____

5. The information gained from this task is______________for measuring your System Knowledge.

   useless--1--------2--------3--------4--------5--------6--useful

   Please describe why you rated the task as you did :______________________________
   ____
   ____

6. Please write your comments here and on the back of this page:

Explaining Troubleshooting Steps Task

(The following questions pertain to the task in which we asked you to go back through the troubleshooting problem and explain why you took the actions you took and what the corresponding equipment results meant to you.)

1. This task was: difficult--1--------2--------3--------4--------5--------6--easy

   Please describe why you rated the task as you did :______________________________
   ____
   ____

2. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop this task seemed:

   different--1--------2--------3--------4--------5--------6--similar

   Please describe why you rated the task as you did :______________________________
   ____
   ____

3. The range of responses that I could use to express myself in this task seemed:

   restricted--1--------2--------3--------4--------5--------6--broad

   Please describe why you rated the task as you did :______________________________
   ____
   ____

4. In comparison to the actual troubleshooting of the RWR Wiring problem in the shop, this task seemed:

   artificial--1--------2--------3--------4--------5--------6--realistic

   Please describe why you rated the task as you did :______________________________
   ____
   ____

5. The information gained from this task is______________for measuring your System Knowledge.

   useless--1--------2--------3--------4--------5--------6--useful

   Please describe why you rated the task as you did :______________________________
   ____
   ____

6. Please write your comments here and on the back of this page:

Appendix C: Questionnaire 2

Task Listing:

<u>Card Task:</u> Generating system components-written on index cards and laid out in front of you.

<u>Ratings Task</u>: Rating the relatedness of pairs of system components, using the computer.

<u>Diagraming Task:</u> Arranging and connecting a set of system components, using index cards and paper arrows.

<u>Think Aloud Task:</u> Talking aloud continuously while troubleshooting the problem.

<u>Troubleshooting Task:</u> Troubleshooting the RWR Wiring problem verbally, specifying your actions with Sgt. Kruse reporting the equipment states.

<u>Explaining Troubleshooting Steps Task:</u> Reviewing actions and equipments states in completed troubleshooting problem.

---

1. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Troubleshooting Task   B. Think Aloud Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar   Extremely similar

2. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Diagramming Task   B. Think Aloud Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar   Extremely similar

3. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Troubleshooting Task   B. Ratings Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar … Extremely similar

4. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Diagramming Task   B. Ratings Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar … Extremely similar

5. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Card Task   B. Explaining Troubleshooting Steps Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar … Extremely similar

6. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

   A. Explaining Troubleshooting Steps Task   B. Ratings Task

   Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

   1--------2--------3--------4--------5--------6--------7--------8
   Not at all similar … Extremely similar

7. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Explaining Troubleshooting Steps Task   B. Diagramming Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar   Extremely similar

8. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Card Task   B. Think Aloud Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar   Extremely similar

9. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Think Aloud Task   B. Explaining Troubleshooting Steps Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar   Extremely similar

10. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Card Task   B. Troubleshooting Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar   Extremely similar

11. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Think Aloud Task B. Ratings Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar — Extremely similar

12. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Card Task B. Diagramming Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar — Extremely similar

13. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Ratings Task B. Card Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar — Extremely similar

14. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Diagramming Task B. Troubleshooting Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar — Extremely similar

15. Please circle the task below which best measures the knowledge necessary for the actual troubleshooting of the RWR Wiring problem in the shop.

A. Troubleshooting Task   B. Explaining Troubleshooting Steps Task

Please rate the extent to which the knowledge measured by the task you circled above is similar to the knowledge needed for the actual troubleshooting of the RWR Wiring problem in the shop.

1--------2--------3--------4--------5--------6--------7--------8

Not at all similar   Extremely similar

16. In the space below, please write your own definition of System Knowledge:

17. We think of System Knowledge as knowledge of the components of a system, how those components fit together, and how they work together. Do you think System Knowledge is important for the actual troubleshooting of the RWR Wiring problem in the shop? Indicate your response on the rating scale below:

1	2	3	4	5
Not at all Necessary	Slightly Important	Important	Extremely Important	Impossible to solve problem without it

Appendix D: Supervisor Rating Form

Rater______________________

Total Active Federal Military Service: ____years, ____months

Total Time in this Career Field: ____years, ____months

IDENTIFICATION OF RATEE

In this survey you are being asked to give

information about the person identified below:

____

In all the sections which follow, any reference to "this person,"

this airman," the ratee," or "the person being rated" means

the person identified above.

Section I: Familiarity with Ratee

For each item below, please read the question and
provide the response that best describes your answer.

1. How long have you served as this person's supervisor? ____ years, ____months.

2. How often do you supervise this person? (check one)

   ___Every Day ___2-3 times per week ___1-3 times/month ___less than once/month

3. How many airmen have you supervised who have the same grade and AFSC as the person being rated? (Counting this airman as "1", please check the appropriate response below.)

   ____ 1 or 2 ____ 3 to 5 ____ 6 to 10 ____ 11 or more

4. To what extent does this airman perform a "mainstream" job within his/her AFSC? In other words, to what extent does this person's job consist of tasks that are typically performed by members of this AFSC? (Check the word below which best completes the following statement)

   His/Her job contains __________ typical tasks.

   ____ no ____ few ____ some ____ mostly

5. Please rate this airman according to his/her general job knowledge:

   Novice--1-----2-----3-----4-----5-----6----7--Expert

Section II: Job Performance

In this section of the survey you will be rating the airman's job performance. You will be completing these ratings in the context of a specific troubleshooting problem. Please assume that the airman has been asked to troubleshoot the following problem within the TEWS system:

The RWR is inoperative. The BIT light is on. The TEWS display is blank.

(Connector 65P-D002H on LRU3 has one of the wires in the twisted pair shorted to the shield in the connector. This wire pair lead to the TEWS Display Unit (65P-J009A) and provides the data in the path.)

Please rate the airman's job performance in reference to this specific problem. Each statement below expresses some aspect of job performance. Use the following rating scale to indicate the extent to which you agree or disagree that the statement is an accurate description of the airman being rated. Remember, all ratings should be completed in the context of the above troubleshooting problem.

RATING SCALE FOR JOB PERFORMANCE

7 = Strongly Agree
6 = Agree
5 = Slightly Agree
4 = Neither Agree nor Disagree
3 = Slightly Disagree
2 = Disagree
1 = Strongly Disagree

1. Would need very little supervision in performing this assignment .... _______
2. Would not likely need to ask others for help in performing this assignment .... _______
3. Would meet local demands for speed and accuracy in carrying out this assignment .... _______
4. Would likely serve as consultant to other workers carrying out this assignment.......... _______
5. Would be capable of performing jobs other than carrying out this assignment .... _______
6. Would carry out this assignment to the best of his/her ability .... _______
7. Would cooperate with supervisors & co-workers if assignment called for teamwork.. _______
8. Has a strong sense of responsibility to the unit .... _______
9. Displays willingness to do more than the required amount of work .... _______
10. Adjusts quickly and effectively to changing work situations .... _______

Section III: Job Knowledge

In this section of the survey you will be rating the airman's job knowledge. You will be completing these ratings in the context of the same troubleshooting problem. Again, please assume that the airman has been asked to troubleshoot the following problem:

> The RWR is inoperative. The BIT light is on. The TEWS display is blank. (Connector 65P-D002H on LRU3 has one of the wires in the twisted pair shorted to the shield in the connector. This wire pair lead to the TEWS Display Unit (65P-J009A) and provides the data in the path.)

Please rate the airman's job knowledge in reference to <u>this specific problem</u>. Use the following rating scale to indicate how the airman compares with other airmen in this career field who are at the same grade level. Remember, <u>all ratings should be completed in the context of the above troubleshooting problem</u>.

<u>RATING SCALE FOR JOB KNOWLEDGE</u>

7 = Very Much Above Average
6 = Above Average
5 = Slightly Above Average
4 = Average Knowledge Compared to other Airmen
3 = Slightly Below Average
2 = Below Average
1 = Very Much Below Average
X = Not Applicable to This Specialty
? = I don't know this about the airman being rated

1. Airman's knowledge of Career Development Course (CDC) material needed for this task........ _______
2. Airman's knowledge of essential technical procedures used in this task .... _______
3. Airman's knowledge of the mission of the unit/organization .... _______
4. Airman's knowledge of the more difficult tasks in the specialty .... _______
5. Airman's knowledge of the technical requirements of the specialty ouside his/her present job. .... _______
6. Airman's knowledge of safety procedures pertinent to this task .... _______
7. Airman's knowledge of time-saving techniques pertinent to this task .... _______
8. Airman's knowledge of required forms pertinent to this task .... _______
9. Airman's knowledge of technical reference materials such as operating manuals, Technical Orders (TOs), or standard reference books.pertinent to this task .... _______
10. Airman's knowledge of how to locate other people with specialized expertise pertinent to this task........ _______
11. Airman's knowledge of the proper technical terminology needed to discuss objects, methods, and goals pertinent to this task........ _______

RATING SCALE FOR JOB KNOWLEDGE

7 = Very Much Above Average
6 = Above Average
5 = Slightly Above Average
4 = Average Knowledge Compared to other Airmen
3 = Slightly Below Average
2 = Below Average
1 = Very Much Below Average
X = Not Applicable to This Specialty
? = I don't know this about the airman being rated

12. Airman's knowledge of the underlying principles, ideas, or concepts pertinent to this task .... _______
13. Airman's knowledge of what can go wrong in completing this task and how to avoid these problems .... _______
14. Airman's knowledge of how to cope with unexpected problems in technical assignments .... _______
15. Airman's knowledge of the tools and equipments pertinent to this task .... _______

Appendix E: Verbal Protocol and Action Coding Scheme

RWR Wiring--Broken Wire Problem Taxonomy

+: Verbalization units used for coding the problem

#: Action units used for coding the problem

Debrief Questions

1.0 Ask pilot: Have an ECS light? #

2.0 Ask pilot: Any indications on MCCP when turned RWR on? #

3.0 Ask pilot: Any indications on TEWS scope when turned RWR on? #

4.0 Ask pilot: Did malfunction exist during entire flight? #

5.0 Ask pilot: Did TEWS scope flicker on at any time? #

6.0 Ask pilot: Everything O.K. before flight? #

7.0 Ask pilot: Cycle RWR off/on? #

8.0 Ask pilot: Check ASP panel? #

9.0 Ask pilot: Do you have an AI/SAM light now? #

10.0 Ask pilot: Did you ever have an AI/SAM light? #

11.0 Ask pilot: Was the AI/SAM light on hard? #

12.0 Ask pilot: Did you have a flashing AI/SAM light? #

13.0 Ask pilot: Has RWR been on for awhile and if so is it still on? #

14.0 Ask pilot: Did you do a BIT check? #

15.0 Ask pilot: Was the BIT light on steady the whole time? #

16.0 Ask pilot: Did you have a BIT light? #

17.0 Ask pilot: Did you try to defeat the tones? #

18.0 What is my fault reporting code? #

Visual Inspections

19.0 Check ASP panel #

20.0 Check LRU Latches

- 20.1 Check problem-appropriate LRU latches #
  - 20.1.1 LRU2 latches (under Door 6-R)
    - 20.1.1.1 Prior to letting RWR run
    - 20.1.1.2 After letting RWR run for approx 15-30 min.
  - 20.1.2 LRU3 latches (under Door 6-R)
    - 20.1.2.1 Prior to letting RWR run
    - 20.1.2.2 After letting RWR run for approx 15-30 min.

20.1.3 LRU9 latches

20.1.3.1 Prior to letting RWR run

20.1.3.2 After letting RWR run for approx 15-30 min.

20.2 Check problem-inappropriate LRU latches #

20.2.1 LRU6 latches

20.2.1.1 Prior to letting RWR run

20.2.1.2 After letting RWR run for approx 15-30 min.

21.0 Check RWR circuit breakers #

22.0 Check LRU Connections

22.1 Check problem-inappropriate LRU Connections #

22.1.1 LRU2 connections

22.1.1.1 Cannon plugs

22.1.1.2 Pins

22.1.1.3 ARF connectors

22.1.2 LRU3 connections

22.1.2.1 Cannon plugs

22.1.2.2 Pins

22.1.2.3 ARF connectors

22.1.3 LRU9 connections

22.1.3.1 Cannon plugs

22.1.3.2 Pins

22.1.3.3 ARF connectors

22.2 Check problem-inappropriate LRU Connections #

22.2.1 LRU6 connections

22.2.1.1 Cannon plugs

22.2.1.2 Pins

22.2.1.3 ARF connectors

23.0 Visually check aircraft wiring #

23.1 Pull cannon plug off & visually check wires 65P-D002H--65P-D009A

23.2 Visually check for chaff

23.2.1 65P-D002H--65P-D009A

23.2.1.1 from LRU3 end

23.2.1.2 from LRU9 end

23.3 Visually check pins

24.0 Check LRU2 fuses #

25.0 Check LRU ETI meters

25.1 Easy Check: ETI meter #

25.1.1 LRU2 ETI meter

25.1.2 LRU3 ETI meter

25.2 Difficult Check: ETI meter #

25.2.1 LRU6 ETI meter

25.2.2 LRU9 ETI meter

26.0 Turn control panel off/on; Cycle RWR off/on. #

27.0 Turn RWR on and check if BIT light flashes. #

28.0 Is the RWR on/off switch in "on" position? #

29.0 Turn RWR on--does AI/SAM light flash? #

30.0 Is A/I SAM light on or off, steady? #

31.0 Adjust TEWS scope intensity to full bright. #

32.0 Did EWW light go off when turned on? #

33.0 Do fault indicators reset? #

73.0 Turn RWR off--Do latches remain latched?#

Built-in Tests (BIT)

34.0 RWR BIT check (duplicate symptoms?) #

35.0 RWR BIT check--let RWR run for > 5 minutes (duplicate symptoms?) #

36.0 Other system BIT #

36.1 IBS (Interference Blanker) BIT

36.2 CC (Central computer) BIT

Reprogram LRU

37.0 Reprogram LRU3. #

Audio Inspections

38.0 Check: Getting audio tones? #

Aircraft History

39.0 Check forms: Is this a repeat failure? #

Swaps

40.0 Swap LRU:

40.1 Swap Problem-appropriate LRU #

40.1.1 R/R LRU2

40.1.2 R/R LRU3

40.1.3 R/R LRU9

40.2 Swap Problem-inappropriate LRU #

40.2.1 R/R LRU6

40.2.2 R/R LRU10

40.2.3 R/R LRU11

41.0 R/R ASP panel #

42.0 R/R (or repair) video cable #

43.0 R/R wire

43.1 65P-J009C #

43.2 65P-J009A #

44.0 R/R BIT Control Panel #

45.0 R/R connector

45.1 On LRU9 side. #

45.2 On LRU3 side. #

Measurements: Wire Continuity

46.0 Measure wire continuity LRU2--LRU3 #

46.1 65P-D001G (LRU2) to 65P-D002L (LRU3)

46.1.1 all pins to all pins

46.2.2 pin 67 and pin 67.

46.2.3 pin 127 and pin 127.

46.2 from 65P-D001E (LRU2) to 65P-D002D (LRU3)

46.2.1 pin 27 to pin 27

46.2.2 pin 28 to pin 28

46.2.3 pin 30 to pin 30

46.2.4 pin 31 to pin 31

46.2.5 pin 127 (LRU2) to pin 126 (LRU3).

47.0 Measure wire continuity LRU6--LRU3 #

47.1 65P-T005G (LRU6) to 65P-D002A (LRU3)

47.2 5J7 (LRU6) to 2J1 (LRU3)

47.3 5J11 (LRU6) to 2J5 (LRU3)

47.4 5J6 (LRU6) to 2J7 (LRU3)

48.0 Measure wire continuity LRU6--LRU2 #

48.1 5J1 (LRU6) to 1J3 (LRU2)

49.0 Measure wire continuity LRU2--LRU9 #

49.1 All wires, all voltages from LRU2 to LRU9.

49.2 65P-J001B to 65P-D009B

49.2.1 pin 53 to pin 53

49.2.2 pin 25 to pin 25 for 28 Volts

49.2.3 pin 54 to pin 54 for 115 Volts AC

49.2.4 pin 19 to pin 19 for -25 Volts.

49.2.5 pin 17 to pin 17 for 100 Volts.

49.2.6 pin 23 to pin 23 for -5 Volts.

49.2.7 pin 27 to pin 27.

49.2.8 pin 28 to pin 28.

49.2.9 pin 29 to pin 29.

49.2.10 pin 30 to pin 30.

49.2.11 pin 31 to pin 31.

49.3 1J2 (LRU2) to 9J2 (LRU9)

49.3.1 end-to-end

49.3.2 shorts between wires

50.0 Measure wire continuity Circuit Breaker Panel--LRU2 #

50.1 Circuit Breaker Panel to LRU2, 1J1

50.1.1 end-to-end

50.1.2 each wire to A/C ground

51.0 Measure wire continuity--Lights Test Relay Panel #

51.1 Through Lights Test Relay Panel

51.2 pins 4 to 27

52.0 Measure wire continuity--Miscellaneous #

52.1 Measure wire continuity from 65P-D002B (LRU3), pins 11 and 13 to pin 25 andA/C grnd.

52.2 Place BCP selection switch to RWR and press intitiate.

52.3 TEWS circuitry

52.4 Power wires coming from LRU3 (?)

53.0 Measure wire continuity LRU3--LRU9

53.1 Between 65P-D002H and 65P-J009A

53.1.1 Center conductor to shield #

53.1.1.1 Center conductor to shield, someone at other end shorting shield.

53.1.1.2 Pin 1 to shield #

53.1.1.3 Pin 2 to shield #

53.1.2 Pin 1 to Pin 2 #

53.1.2.1 on LRU3 side

53.1.2.2 on LRU9 side

53.1.2.3 Jump from center conductor (pin 1) to outside conductor (pin 2).

53.1.3 Center pin to A/C ground #

53.1.3.1 Pin 1 to A/C ground

53.1.3.1.1 With ground (shorting wire 1 to ground)

53.1.3.1.2 With no ground

53.1.3.1.3 Pin 1 to ground at each end

53.1.3.2 Pin 2 to A/C ground.

53.1.3.2.1 With ground (shorting wire 1 to ground)

53.1.3.2.2 With no ground.

53.1.3.2.3 Pin 2 to ground at each end

53.1.4 Center pins: End-to-end #

53.1.4.1 Pin 1 to Pin 1 (end-to-end)

53.1.4.2 Pin 2 to Pin 2 (end-to-end)

53.1.5 Shield to A/C ground #

53.1.5.1 With ground (shorting wire 1 to ground)

53.1.5.2 Shield to A/C ground, no ground

53.1.5.3 Pin 1 to A/C ground

53.1.5.3.1 on LRU3 side

53.1.5.3.2 on LRU9 side

53.1.6 Shield to shield #

53.1.7 Remove LRU3 connector and shoot the bare pins, bare wires. #

53.2 Between 65P-D002J and 65P-J009C (off the active path) #

53.2.1 Pin to pin

53.2.1.1 Pin 3 to Pin 4

53.2.1.2 On 65P-J009C: jumper both pins from LRU3 back to the LRU9.

53.2.1.3 Pin to pin (put jumper lead on center conductor to A/C grnd and check 2J end of cable.

53.2.2 Center conductor to shield.

53.2.2.1 Pin 3 to shield

53.2.2.2 Pin 4 to shield

53.2.3 Center pins: End-to-end

53.2.3.1 Pin 3 to Pin 3

53.2.3.2 Pin 4 to Pin 4

53.2.4 Shield to shield

53.2.5 Center conductor to ground

53.2.5.1 Pin 3 to ground at each end

53.2.5.2 Pin 4 to ground at each end

53.3 Between 65P-D001F and 65P-J002P (off the active path) #

53.3.1 Pin 4 to pin 4.

53.3.2 Pin 5 to pin 5.

53.3.3 Pin 1 to pin 1.

53.3.4 Pin 2 to pin 2.

54.0 Measure wire continuity TEWS Control Panel--LRU3 #

54.1 pin 1 to pin 1

55.0 Measure wire continuity TEWS Control Panel--Miscellaneous #

55.1 from 65P-J010 (TEWS CP), pin 2 to Avionics Protection Relay #5, pin 4

55.3 from 65P-J010 (TEWS CP), pin 2 to 52P-J083B (Light Test Relay Panel), pin 4

56.0 Measure wire continuity Light Test Relay Panel--LRU3 #

56.1 pin 27 to pin 2

57.0 Measure wire continuity Data Processor--Input Data Link #

Measurements: Voltage/Power

58.0 Measure Voltage/Power to the LRU2 #

58.1 28 Volts DC on 1J2, pin 25 to 32

58.2 Check pin 5--getting 28 Volts DC to LRU2?

59.0 Measure Voltage/Power--General #

59.1 Is the system getting power?

Verbalizations

60.0 Action interpretations

60.1 Power +

60.2 Duplicate fault +

60.3 Connections +

60.4 Ease of Action +

60.5 T.O. Says... +

60.6 Previous result +

60.7 Latch +

60.8 Information Flow +

60.9 Display +

60.10 Tones +

60.11 Shorts +

60.12 Continuity +
60.13 AI/SAM Light +
60.14 BIT light +
60.15 Solve problem/solution +
60.16 Bad parts +
60.17 Miscellaneous/Frustration +

61.0 Result interpretation +
62.0 Elimination +
63.0 Elimination Justification +
64.0 Plan/prepare for test/check +
65.0 T.O. Search/interpretation +